DANIEL J. LEVITIN

LA MENTIRA COMO ARMA

CÓMO PENSAR CRÍTICAMENTE EN LA ERA DE LA POSVERDAD

Alianza editorial

Daniel J. Levitin

LA MENTIRA COMO ARMA

CÓMO PENSAR CRÍTICAMENTE EN LA ERA DE LA POSVERDAD

Traducción de Jesús Martín Cordero

Alianza editorial

A mi hermana Shari, cuya inquisitiva mente
me hizo un pensador mejor

Índice

INTRODUCCIÓN

PENSAR, CRÍTICAMENTE

Voy a comenzar afirmando dos cosas que seguramente van a enfadar mucho a algunas personas. Primera, la lengua que utilizamos ha comenzado a empañar la relación entre los hechos y la fantasía. Segunda, esto es una secuela peligrosa de la falta de educación en nuestro país y de cómo ha afectado a toda una generación de ciudadanos. Estos dos hechos han provocado que las mentiras proliferen en nuestra cultura de un modo nunca visto. Ha sido posible convertir las mentiras en armas de modo que socaven subrepticiamente nuestra capacidad para adoptar decisiones correctas para nosotros mismos y para nuestros conciudadanos.

¿Qué ha pasado con nuestra lengua? La palabra del año del Diccionario Oxford de 2016 fue *posverdad,* definida como un calificativo «relacionado con o que denota circunstancias en las que los hechos objetivos resultan menos influyentes en la formación de la opinión pública que apelar a las emociones y las creencias personales». Se seleccionó esa palabra porque su uso se había disparado en ese año. Creo que deberíamos retornar al uso de la vieja y llana «verdad» y deberíamos hacerlo rápido. Y necesitamos rechazar la idea de que la verdad ya no existe.

Estamos siendo demasiado puntillosos cuando nos referimos a las falsedades. Tal vez con el ánimo de evitar enfrentamientos personales, al esforzarnos por «llevarnos bien», hemos comenzado a utilizar eufemismos para referirnos a cosas que son locuras absurdas. La mentira de que la pizzería Comet Ping Pong, de Washington, desarrollaba un negocio de esclavitud sexual, lanzada por Hillary Clinton, llevó a Edgar M. Welch, de veintiocho años, residente en Salisbury, Carolina del Norte, a conducir quinientos sesenta kilómetros desde su casa hasta

Washington y disparar su arma semiautomática dentro de la pizzería el domingo 4 de diciembre de 2016 (pocos días después de que «posverdad» se convirtiera en la palabra del año). El *Daily News* de Nueva York afirmó que esta mentira era una «teoría marginal». Por cierto, una teoría no es meramente una idea, es una idea basada en una evaluación cuidadosa de las pruebas. Y no basta cualquier prueba, deben ser pruebas pertinentes para el asunto que se trata y obtenidas de forma rigurosa y no sesgada.

Otros eufemismos para referirse a las mentiras son contraconocimiento, medias verdades, concepciones radicales, verdad alternativa, teorías de la conspiración y, más recientemente, noticias falsas *(fake news)*.

La expresión «noticia falsa» suena demasiado trivial, un poco como cuando un niño simula estar enfermo para no ir a un examen. Estos eufemismos ocultan el hecho de que la historia de la esclavitud sexual es una mentira pura y dura, quienes la escribieron sabían que no era verdad. Una historia no tiene dos versiones cuando una de ellas es una mentira. Los periodistas y el resto de nosotros debemos dejar de adjudicar la misma atención a asuntos que carecen de apoyo en los hechos. Una historia tiene dos versiones cuando existen pruebas favorables a dos interpretaciones; en ese caso, las personas razonables pueden estar en desacuerdo sobre la interpretación de las pruebas y sobre qué conclusión extraer de ellas. Cada uno tiene derecho a formar sus propias opiniones, pero no tenemos derecho a disponer de los hechos. Las mentiras son la ausencia de hechos y en muchos casos se oponen directamente a ellos.

La verdad sí importa. Una era de la posverdad es una era de irracionalidad deliberada, que se opone a todos los grandes avances de la humanidad. Es posible que los periodistas no quieran llamar a las «noticias falsas» lo que son, mentiras puras y duras, porque no desean ofender a los mentirosos, ¡pero hay que ofenderlos! Expongámoslos a la luz.

Quizá sea preferible expresarlo del siguiente modo: ¿qué ha sucedido en nuestro sistema y nuestras instituciones educativas hasta llegar a esta era de la posverdad? El promedio de libros que leen los estudiantes se reduce cada año tras el *segundo*

curso de primaria[1]. Hace ya quince años que el Ministerio de Educación de los Estados Unidos observó que más de una quinta parte de los estadounidenses no son ni siquiera capaces de buscar una información[2] en un texto o de «realizar inferencias de bajo nivel utilizando texto impreso». Según parece, no hemos conseguido enseñar a nuestros niños qué es una prueba y cómo evaluarla. Esto debería escandalizarnos. Edgar Welch, el tirador de la pizzería Comet Ping Pong, informó a las autoridades que estaba «investigando» la teoría de la conspiración tras leer sobre ella en la red. Disponemos de una potente infraestructura de la información que puede ser beneficiosa, pero también puede ser dañina y cada uno de nosotros debe saber cómo diferenciar lo uno de lo otro.

Puede que Welch creyera de un modo u otro que investigaba, pero no existe la menor prueba de que realizase ninguna investigación. Lo que parece es que este ciudadano ignorante no conoce en qué consiste recopilar y evaluar pruebas. En este caso podríamos buscar algún vínculo entre Hillary Clinton y el restaurante, buscar conductas de Clinton que indiquen algún interés por dirigir un circuito de prostitución o incluso alguna razón por la que pudiera beneficiarse de tal hecho (con seguridad el motivo no puede haber sido financiero, dado el reciente revuelo que han provocado sus tarifas para impartir conferencias). Cabe que haya observado a prostitutas infantiles y a sus clientes entrando o saliendo del restaurante. O, careciendo de la formación y la disposición necesarias para realizar su propia investigación, podría apoyarse en los profesionales, informándose sobre lo que los periodistas de investigación hayan podido escribir sobre esta historia. El hecho de que ningún periodista profesional otorgue el menor crédito al asunto ya dice mucho. Me consta que hay personas que piensan que los periodistas son corruptos y están dirigidos por el Gobierno. El Departamento de Estadísticas Laborales de Estados Unidos informa que hay 45.790 reporteros y corresponsales. La Sociedad Americana de Editores de Noticias, una asociación comercial independiente, estima que hay 32.900 periodistas[3] que trabajan para casi 1.400

periódicos en los Estados Unidos[4]. Es perfectamente posible que algunos periodistas sean corruptos, pero, por muchos que haya, es muy improbable que todos ellos lo sean.

Facebook se esfuerza por cumplir su responsabilidad social como fuente de información «facilitando a sus 1.800 millones de miembros que denuncien las noticias falsas»[5]. En otras palabras, que llamen mentiras a las mentiras. Tal vez otros medios sociales de la red puedan adoptar una disposición más documentada en el futuro. Como mínimo podemos esperar que disminuya su papel en la constitución de las mentiras como armas.

Muchas organizaciones han indagado dónde se originó la historia de la esclavitud sexual en la pizzería. La NBC informó de la existencia de una activa comunidad de creadores de «noticias falsas»[6] en la ciudad de Veles, en Macedonia, que bien podría haber sido la fuente de origen. Esta región formaba parte de la Yugoslavia comunista hasta 1991. *Buzzfeed* y *The Guardian* encontraron más de 100 dominios de noticias falsas originarios de allí. Algunos jóvenes de Veles, gente que no tiene ninguna relación con los partidos políticos de Estados Unidos, elaboran historias basadas en mentiras a fin de obtener ingresos por click de los anunciantes en plataformas como Facebook. Los adolescentes pueden llegar a ganar decenas de miles de dólares en ciudades que ofrecen pocas oportunidades económicas. ¿Hemos de culparlos por el tiroteo de la pizzería? ¿O a las redes sociales? ¿O al sistema educativo estadounidense por haber creado ciudadanos satisfechos de sí mismos que no se toman el trabajo de pensar con un poco de atención sobre las declaraciones que se encuentran diariamente?

Quizá podría objetarse y decir «pero mi trabajo no consiste en evaluar críticamente estadísticas. Los periódicos, los blogueros, el gobierno, la Wikipedia, etc., son quienes tienen esa responsabilidad». Cierto, deberían hacerlo, pero no siempre lo hacen y cada vez les resulta más difícil, ya que la cantidad de mentiras prolifera más rápidamente de lo que cuesta desmentirlas. Es como uno de esos juegos de ordenador en los que siempre surgen bichos nuevos a los que hay que eliminar. La

historia del «Pizzagate» recibió más de un millón de visualizaciones, en tanto que su desmentido en *Snopes*[7] recibió menos de 35.000. Felizmente tenemos libertad de prensa, a lo largo de la historia otras naciones han estado mucho peor, pero nunca deberíamos dar por hecha la libertad y la integridad de los medios. Los periodistas y las empresas que les pagan seguirán ayudándonos a identificar las mentiras y a desactivarlas, pero no pueden hacer eso solos, las mentiras ganarán si tenemos un público crédulo y poco instruido que las consume.

Está claro que la mayoría de nosotros no creemos que Hillary Clinton dirigiera un circuito de esclavitud sexual en una pizzería de Washington, pero este libro no se limita a abordar tales absurdos. ¿Realmente necesitamos ese nuevo fármaco o estamos siendo influidos por una campaña de marketing multimillonaria y basada en datos parciales y sesgados? ¿Cómo podemos saber si un famoso que está siendo juzgado es realmente culpable? ¿Cómo evaluamos una u otra inversión, o un conjunto contradictorio de encuestas electorales? ¿Qué cosas están fuera del alcance de nuestro conocimiento porque no disponemos de suficiente información?

La mejor defensa contra los embusteros taimados, la defensa más fiable, es que todos nosotros nos convirtamos en pensadores críticos. No hemos logrado enseñar a nuestros hijos a contrarrestar la tendencia evolucionista a la credulidad. Somos una especie social y tendemos a creernos lo que los demás nos dicen. Nuestros cerebros son máquinas magníficas de contar historias y fabular: dada una premisa extravagante, somos capaces de generar explicaciones rocambolescas para justificar su veracidad. Justamente ésa es la diferencia entre pensamiento creativo y pensamiento crítico, entre las mentiras y la verdad: la verdad cuenta con pruebas objetivas que la apoyan. Algunas declaraciones *podrían* ser verdad, pero las declaraciones verídicas *son* verdad.

Una investigación de la Universidad de Stanford sobre razonamiento cívico en línea evaluó a 7.800 estudiantes, desde la enseñanza media hasta la universidad, a lo largo de dieciocho meses, finalizando en junio de 2016. Los investigadores indican

«una regularidad sorprendente y lamentable. En conjunto, la capacidad de los jóvenes para razonar sobre la información de internet puede resumirse con una palabra: desoladora». Los jóvenes resultaron muy poco competentes para diferenciar la información de alta calidad de las mentiras. Necesitamos empezar a enseñarles a hacerlo y necesitamos empezar ya. En tanto que comenzamos a enseñarles, al resto de la población le vendría bien un cursillo de refuerzo. Afortunadamente, el razonamiento basado en las pruebas no está fuera del alcance de la mayoría de los jóvenes de doce años, siempre que se les muestre cómo hacerlo.

Hay mucha gente que afirma que el «Pizzagate» fue un resultado directo de las noticias falsas, pero llamemos a las cosas por su nombre: mentiras. No hay «noticias» en las noticias falsas. Creer en algunas mentiras puede resultar inocuo, como la creencia en Papá Noel o en que estos vaqueros nuevos me hacen parecer delgado. Lo que convierte a las mentiras en *armas* no son los medios de comunicación o Facebook. El peligro reside en la intensidad de la creencia, la confianza incuestionable de que se trata de una verdad.

El pensamiento crítico nos permite retroceder un paso para evaluar los hechos y establecer conclusiones basadas en las pruebas. Lo que condujo a Welch a disparar su arma en una pizzería de Washington fue su completa incapacidad para comprender que su creencia podía estar equivocada. El componente más importante del mejor pensamiento crítico del que hoy carece nuestra sociedad es la humildad. Se trata de una noción simple pero profunda: si nos damos cuenta de que no lo sabemos todo, podemos aprender. Si creemos que nos lo sabemos todo, aprender resulta imposible. De algún modo nuestro sistema educativo y nuestra confianza en internet han conducido a una generación de chavales a no saber que no saben. Si fuéramos capaces de aceptar esa verdad, podríamos educar la mente de los estadounidenses, restaurar el civismo y desarmar toda una plétora de mentiras armadas que amenazan nuestro mundo. Es la única vía por la que la democracia podrá prosperar.

Tres tipos de defensa estratégica

Comencé a escribir este libro en 2001, mientras ofrecía un curso universitario sobre pensamiento crítico. Me dediqué en serio al libro durante 2014-2016 y lo publiqué con una Introducción diferente y bajo el título de *A Field Guide to Lies* (Guía de campo sobre las mentiras). Desde entonces la peligrosidad y el alcance de las mentiras han llegado a ser abrumadores. Ya no se limitan a ser asuntos sobre los que la gente hace sarcasmos o que provocan risitas; se han convertido en armas. Este peligro podría empeorar, podría provocar problemas a los que no hemos hecho frente durante generaciones. O podría desvanecerse sin llegar a consecuencias drásticas. De cualquier modo, las herramientas que ofrezco aquí son las mismas que en la primera edición; son herramientas necesarias, independientemente de los vientos que soplen en los ámbitos político, social y económico.

Parte del problema tiene que ver con las fuentes. En los viejos tiempos, los libros sobre hechos reales y los artículos de periódico parecían auténticos si se comparaban con cualquier tocho que un chiflado hubiera editado en una imprenta doméstica en su sótano. Sin duda, Internet ha cambiado todo esto. Una web excéntrica puede parecer tan auténtica como una basada en hechos y acreditada, más adelante ofrezco algunos ejemplos. En Internet la desinformación está diabólicamente mezclada con la información real, haciendo que sea difícil diferenciar una de otra. La desinformación es promiscua, aparece vinculada a personas de todas las clases sociales y niveles educativos y surge en lugares inesperados. Se propaga de una persona a otra a medida que Twitter, Facebook, Snapchat, Instagram, Tumblr y otros medios sociales la difunden por el planeta; la desinformación puede arraigar y acabar siendo muy conocida y, de repente, grupos enteros de personas creen cosas que no son como piensan.

Este es un libro sobre cómo identificar problemas en los hechos que uno se encuentra, problemas que pueden llevarnos a extraer conclusiones equivocadas. A veces las personas que nos proporcionan los hechos esperan que saquemos conclusiones

equivocadas; a veces ellos mismos no conocen la diferencia. Hoy día la información está disponible de modo prácticamente instantáneo, los líderes nacionales aparecen en los medios sociales de Internet, diariamente nos llaman la atención «noticias urgentes», incluso cada hora, pero ¿cuándo tenemos tiempo para determinar si una información nueva llega acompañada de pseudohechos, distorsiones o mentiras puras y duras? Todos necesitamos estrategias eficaces para evaluar si lo que nos están contando es confiable.

Hemos producido más información en los últimos cinco años que en toda la historia humana previa. Junto a cosas que son verdad hay una enorme cantidad de asuntos que no lo son, en sitios web, vídeos, libros y en los medios sociales. No se trata meramente de un problema nuevo. La desinformación ha sido un rasgo de la vida humana[8] durante miles de años y está documentada en los tiempos bíblicos y en la Grecia clásica. El problema singular al que hacemos frente hoy es que la desinformación ha proliferado y las mentiras pueden convertirse en armas para alcanzar fines sociales y políticos contra los que estaríamos protegidos si no fuera así.

En los capítulos siguientes he agrupado estas estrategias por categorías. La primera parte de este libro trata de la desinformación numérica. Muestra cómo las estadísticas y los gráficos mal manejados pueden generar perspectivas muy distorsionadas, manipuladas, y conducirnos a conclusiones erróneas (y a tomar decisiones equivocadas). La segunda parte del libro investiga las argumentaciones incorrectas, mostrando cuán fácil es ser persuasivo, contar historias que se desvían de los hechos de modo atractivo pero erróneo. En el trascurso del texto veremos qué medidas podemos tomar para evaluar mejor las noticias, los anuncios, los informes. La última parte del libro revela qué es lo que fundamenta nuestra capacidad para determinar si algo es verdadero o falso: el método científico. Esta es la mejor herramienta inventada para descubrir los misterios más desafiantes y remonta sus raíces a algunos de los mayores pensadores de la historia humana, a figuras como Aristóteles, Bacon, Galileo, Descartes, Semmelweis y Popper. Esa última

parte del libro trata los límites de lo que podemos y no podemos conocer, incluyendo lo que sabemos hoy día y lo que todavía no conocemos. Ofrezco varios estudios de caso para demostrar las aplicaciones del pensamiento lógico a contextos muy diversos, desde el testimonio ante los tribunales hasta la toma de decisiones en medicina, la magia, la física moderna y las teorías de la conspiración.

El pensamiento crítico no implica menospreciar todo lo demás; supone que intentemos diferenciar entre las afirmaciones basadas en los hechos y aquellas que no lo están.

Los creyentes fervorosos lo tienen fácil para mentir con las estadísticas y los gráficos, porque saben que la mayoría de las personas piensan que llevará demasiado tiempo mirar bajo la manta y comprobar cómo lo hacen. Tal vez crean que no son lo suficientemente inteligentes. Pero lo cierto es que cualquiera puede hacerlo y una vez que se dispone de algunos principios básicos, los gráficos revelan rápidamente su elegancia o sus distorsiones.

Considérese la estadística que cité antes sobre el descenso gradual, año a año, de la cantidad de libros que leen los estudiantes desde segundo de primaria. La consecuencia es que nuestro sistema educativo no funciona bien, los niños no desarrollan buenos hábitos de lectura, no se interesan por mejorar, no están comprometidos intelectualmente. Pues bien, detengámonos a pensar y preguntémonos: ¿la *cantidad de libros* es la medida adecuada para alcanzar conclusiones como esta? Habitualmente los alumnos de segundo de primaria leen libros muy breves y la longitud de estos va aumentando con la edad. Los niños pueden leer *El señor de las moscas* (200 páginas) y, cuando llegan a la universidad, *Guerra y paz* (1.225 páginas). Tal vez debiéramos considerar la cantidad de páginas leídas o la cantidad de tiempo que se emplea leyendo. Al llegar a la universidad y en muchas profesiones como el derecho, la administración, las finanzas y la ciencia, las personas pueden leer menos libros, pero sí que leen artículos cada vez más rigurosos y eruditos. Si un funcionario de la Administración no leyera *ningún* libro, pero empleara su tiempo leyendo la Constitución, la

legislación, los informes de inteligencia, los periódicos y revistas, ¿podríamos decir que no está intelectualmente comprometido? El mero hecho de citar una estadística no significa que esta sea relevante para lo que se está discutiendo. Es más, el estudio que cité parece haber sido elaborado por una empresa que diseña y vende software para mejorar las competencias de lectura, de modo que se benefician de un informe que ilustra los bajos niveles de lectura. El pensamiento crítico en acción.

Identificar los argumentos erróneos incorporados en las historias nos ayuda a evaluar si una cadena de razonamientos conduce, o no, a una conclusión válida. La infoalfabetización significa ser capaz de reconocer que existen jerarquías en la calidad de las fuentes, que los pseudo-hechos a menudo se enmascaran como hechos y que los sesgos pueden distorsionar la información que hemos de considerar, conduciéndonos a malas decisiones y malos resultados.

A veces las pruebas toman la forma de números y hemos de preguntarnos «¿de dónde salen esos números? ¿cómo se han recopilado?». A veces los números son ridículos, pero es necesario reflexionar para darse cuenta. A veces hay afirmaciones que parecen razonables, pero tienen su origen en una fuente que carece de credibilidad, como cuando una persona que dice haber sido testigo de un delito no estaba presente en realidad. Este libro nos puede ayudar a no creernos muchas cosas que simplemente no son como se cuentan[9]. Y a frenar en seco a los mentirosos.

---

1 http://www.yalsa.ala.org/thehub/2012/04/16/so-how-much-are-kids-really-reading/.

2 http://nces.ed.gov/pubs93/93275.pdf.

3 https://www.bls.gov/oes/current/oes_nat.htm.

4 http://asne.org/content.asp?contentid=415.

5 http://www.nytimes.com/2016/12/15/technology/facebook-fake-news.html.

6 http://www.nbcnews.com/news/world/fake-news-how-parying-macedonian-teen-earns-thousands-publishing-lies-n692451.

7 *N. del T.:* Snopes (https://www.snopes.com/) es una web de verificación de noticias inventadas, mitos, leyendas urbanas, rumores, etc.

8 El propio Abraham proporciona información falsa sobre la identidad de su esposa, Sarah, al rey Abimelech para protegerse. El caballo de Troya era una especie de información falsa que parecía un regalo, pero contenía soldados.

9 Según D. Huff (1954/1993). *How to Lie with Statistics.* Nueva York: W. W. Norton, p. 9. Y, como veremos más adelante, probablemente se hacía eco de Mark Twain, o de Josh Billings, o de Will Rogers, o de quién sabe quién.

PARTE I

EVALUAR LAS CIFRAS

Lo que desconoces no es lo que va a meterte en problemas, es lo que das por sentado pero no lo está.

Mark Twain

VEROSIMILITUD

Al componerse de cifras, las estadísticas nos resultan hechos fríos, sólidos. Parece como si representasen los hechos que nos ofrece la naturaleza y que fuera mera cuestión de encontrarlos. Pero es importante recordar que quienes *elaboran* las estadísticas son personas. Son las personas quienes deciden qué contabilizar, cómo hacerlo y qué parte de las cifras resultantes van a compartir con nosotros[10]. Las estadísticas no son hechos. Son interpretaciones, y vuestra interpretación puede ser igual de buena o mejor que la de la persona que os está informando.

A veces lo que pasa es que las cifras están equivocadas y lo más sencillo es comenzar por realizar algunas pruebas rápidas de verosimilitud. Tras hacerlo, e incluso si las cifras pasan la prueba de verosimilitud, hay tres tipos de errores que pueden conducirnos a creer cosas que no son: cómo se han obtenido esas cifras, cómo han sido interpretadas y cómo se representan gráficamente.

Bien mentalmente, bien en la parte de atrás de un sobre, podemos determinar rápidamente si una afirmación es verosímil (la mayoría de las veces). No aceptéis las afirmaciones a la primera, elaboradlas un poco.

Cuando realizamos pruebas de verosimilitud no nos ocupamos de las cifras exactas. Esto puede parecer contraintuitivo, pero en ese punto la precisión no es lo que importa. Podemos utilizar el sentido común para considerar lo siguiente: si Bert nos dice que una copa de vino de cristal se cayó sobre una alfombra gruesa sin romperse, eso parece verosímil. Si Ernie dice que se cayó desde un edificio de cuarenta pisos y se golpeó contra el suelo sin romperse, eso no es verosímil. Nos lo dicen nuestros conocimientos sobre el mundo real, las observaciones realizadas a lo largo de nuestras vidas. De modo semejante, si alguien nos dice que tiene doscientos años, o que gana a la ruleta

sistemáticamente en Las Vegas o que es capaz de correr a setenta kilómetros por hora, no se trata de afirmaciones verosímiles.

¿Qué pensaríamos de la siguiente afirmación?

> En los treinta y cinco años que han transcurrido desde que las leyes contra la marihuana dejaron de aplicarse en California, la cifra de fumadores de marihuana se ha duplicado cada año.

¿Verosímil? ¿Dónde empezamos? Supongamos que hace treinta y cinco años en California hubiera un único fumador de marihuana, una estimación bastante conservadora (hubo medio millón de detenciones en 1982). Doblar esa cifra cada año durante treinta y cinco años arroja un resultado de más de 17.000 millones, una población mayor que la del planeta. (Intentadlo vosotros mismos y comprobaréis que doblar la cifra cada año durante veintiún años os lleva a más de un millón: 1; 2; 4; 8; 16; 32; 64; 128; 256; 512; 1.024; 2.048; 4.096; 8.192; 16.384; 32.768; 65.536; 131.072; 262.144; 524.288; 1.048.5769). En consecuencia, la afirmación no solo es inverosímil, es imposible. Lamentablemente muchas personas tienen dificultades para razonar con claridad sobre cifras, ya que estas las intimidan. Pero, como veréis, nada de lo que voy a contar requiere más que la aritmética de la enseñanza primaria y adoptar algunos supuestos razonables.

Vamos con otra. Acabáis de empezar a trabajar en telemarketing, un ámbito en el que los agentes de ventas telefonean a los incautos (y sin duda irritados) clientes. Intentando motivaros, vuestro jefe afirma:

> Nuestro mejor vendedor ha hecho 1.000 ventas en un día.

¿Resulta verosímil? Si intentáis marcar un número de teléfono lo más rápido que podáis os costará aproximadamente cinco segundos. Añadid otros cinco segundos hasta que la llamada comienza a sonar. Supongamos que cada una de las llamadas concluye con una venta; obviamente no es un supuesto realista, pero vamos a hacer todas las concesiones para ver si esta

afirmación se sostiene. Imaginemos que son necesarios un mínimo de diez segundos hasta que obtenemos el tono de llamada y descuelgan el teléfono, a partir de ahí otros cuarenta segundos hasta que obtenemos los datos de la tarjeta de crédito del comprador y su dirección. Esto supone una llamada por minuto (5 + 5 +10 + 40 = 60 segundos), o 60 ventas por hora, o 480 ventas en una jornada laboral hiperactiva de ocho horas, sin descansos. Las 1.000 ventas son inverosímiles, incluso partiendo de las estimaciones más optimistas.

Algunas afirmaciones son más difíciles de evaluar. Veamos un titular de la revista *Time* en 2013:

> Hay más personas con móvil que con W.C.[11].

¿Qué significa esto? Podemos considerar la cantidad de personas de los países en desarrollo que carecen de cuarto de baño y la observación de que muchas personas en los países prósperos poseen más de un móvil. La afirmación parece verosímil, pero eso no significa que debamos aceptarla, simplemente no la rechazamos por ridícula; hemos de utilizar otras técnicas para evaluar esta afirmación, pero pasa la prueba de verosimilitud.

A veces no es sencillo evaluar una afirmación sin investigar un poco por nuestra cuenta. Ciertamente, los periódicos y los sitios web deberían hacer esa tarea, pero no siempre lo hacen y así es como se asientan las estadísticas sin fundamento. Una estadística que recibió mucha atención hace unos años indicaba lo siguiente:

> En Estados Unidos cada año mueren de anorexia 150.000 jóvenes y mujeres[12].

Vamos a evaluar la verosimilitud de esta afirmación. Necesitamos indagar un poco. Según los Centros de Control de Enfermedades de los Estados Unidos, la cifra total de muertes por *todas las causas* de jóvenes y mujeres de edades comprendidas entre los quince y los veinticuatro años es de 8.500. Añadamos las mujeres que tienen entre veinticuatro y cuarenta y cuatro y todavía nos

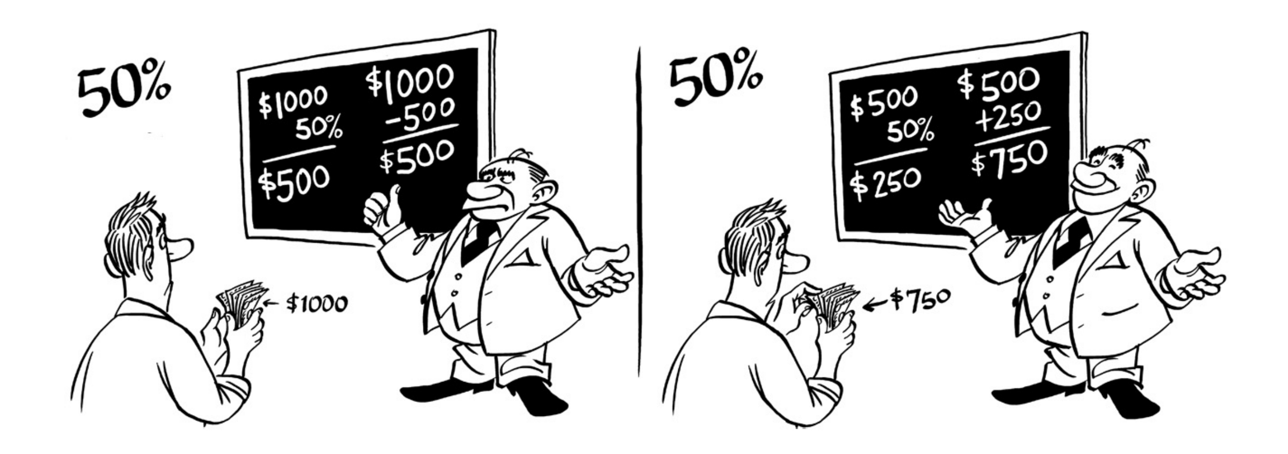
quedamos en 55.000[13]. La cifra de muertes por anorexia al año no puede ser el triple del total de muertes[14].

En un artículo de la revista *Science,* Louis Pollack y Hans Weiss informan que, desde la creación de la empresa Communication Satellite Corp.,

> El precio de las llamadas telefónicas ha disminuido un 12.000%[15].

Si un precio disminuye el 100 por cien, cae a cero (independientemente de cuál fuera el precio inicial). Si un precio disminuye un 200%, alguien os está pagando la misma cantidad que solíais pagar por ese producto. Una disminución del 100 por cien es muy infrecuente; una del 12.000 por cien absolutamente improbable[16]. Un artículo de la revista *Journal of Management Development,* que cuenta con revisión por expertos, anunciaba una reducción del 200 por cien de las reclamaciones de los consumidores[17] tras la puesta en marcha de una nueva estrategia de atención al cliente. El autor, Dan Keppel, incluso tituló su libro *Get What You Pay For: Save 200% on Stocks, Mutual Funds, Every Financial Need* (Obtenga lo que paga: Ahorre un 200% en acciones, fondos de inversión mutuos, en cualquier dificultad financiera). Tiene un título de MBA, debiera hacer mejor las cosas.

Es necesario aplicar los porcentajes a la misma línea de base para poder hacerlos equivalentes. Una reducción salarial del 50%[18] no puede recuperarse incrementando vuestro nuevo salario en un 50%, porque las líneas de base se han alterado. Si cobrabais 1.000 dólares por semana y os lo reducen un 50%, hasta 500, un aumento del 50% solamente significa 750 dólares.



Los porcentajes parecen sencillos e incorruptibles, pero a menudo resultan confusos. Si las tasas de interés suben del 3 al 4%, eso es un incremento de 1 punto porcentual, o del 33% (ya que se realiza un incremento de un 1% sobre una línea de base de 3, de modo que 1/3 = 0,33). Si las tasas de interés bajan del 4 al 3%, es un descenso de 1 punto porcentual, pero no es una caída del 33%, es un descenso del 25% (porque el 1 porcentual de caída se realiza sobre una línea de base de 4). Los investigadores y los periodistas no siempre son escrupulosos a la hora de diferenciar entre puntos porcentuales y porcentajes[19], pero nosotros debiéramos serlo.

El *New York Times* informó del cierre de una fábrica textil de Connecticut y su traslado a Virginia[20] a causa de los altos costes laborales. El *Times* indicaba que estos costes, «los salarios, las indemnizaciones por despido y el seguro de desempleo, son 20 veces más altos en Connecticut que en Virginia». ¿Es esto verosímil? Si fuera cierto, se produciría una migración masiva de empresas de Connecticut a Virginia y ya hubiéramos oído hablar del tema. De hecho, no era cierto y el *Times* fue obligado a publicar una corrección. ¿Cómo sucedió? La periodista se confundió al leer un informe de la empresa. Un coste, el seguro de desempleo, era veinte veces más alto en Connecticut que en Virginia, pero cuando se agregaba a los demás costes, el total de los costes laborales era tan solo 1,3 veces más elevado en Connecticut, no 20 veces más. La reportera carecía de formación en gestión de negocios y no debiéramos suponer que la posea. Pescar estos tipos de error requiere retroceder un paso y pensar

por nuestra cuenta, algo que todos podemos hacer (y que la reportera y sus editores deberían haber hecho).

Nueva Jersey aprobó una legislación que negaba apoyos adicionales[21] a madres que tuvieran hijos mientras aún disfrutaran de prestaciones sociales. Algunos legisladores estimaron que las mujeres en Nueva Jersey tenían hijos simplemente para incrementar la cifra de sus ingresos por prestaciones sociales. Al cabo de dos meses, los legisladores ya afirmaban que la ley del «tapón familiar» era un gran éxito, dado que los nacimientos habían disminuido un 16%[22]. Según el *New York Times:*

> Tras solo dos meses, el Estado ofreció cifras que sugerían que los nacimientos de madres que recibían prestaciones sociales ya habían caído un 16% y los políticos comenzaron a congratularse por su éxito fulgurante.

Obsérvese que no están contando los embarazos, sino los nacimientos. ¿Cuál es el problema? Dado que son necesarios nueve meses de embarazo para que se produzca el nacimiento, cualquier efecto que tenga lugar durante los dos primeros meses no puede atribuirse a la ley, sino que probablemente se deba a las fluctuaciones normales de la tasa de nacimientos (se sabe que las tasas de nacimientos son estacionales).

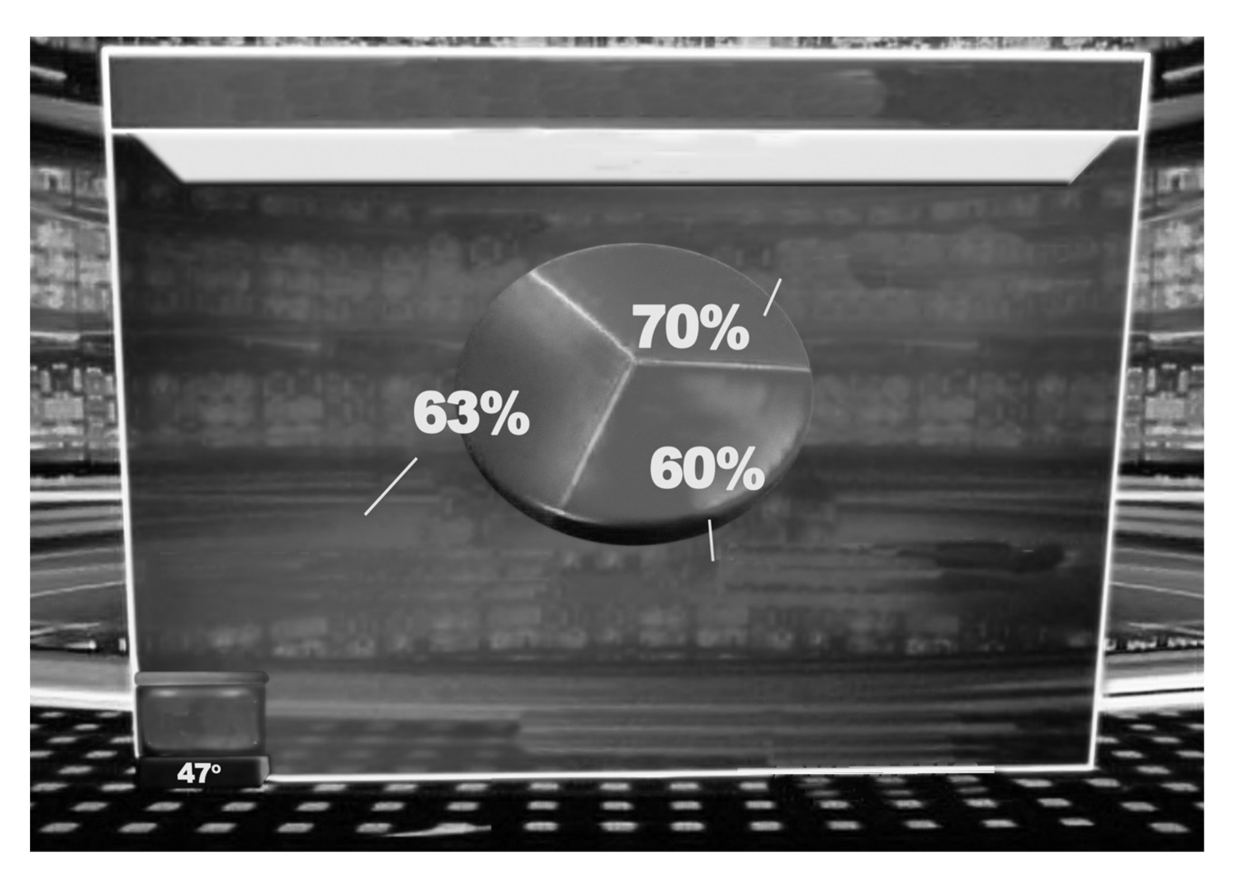
Incluso así, hay otros problemas en este informe que no pueden detectarse mediante pruebas de verosimilitud:

> ... con el paso del tiempo, esa caída del 16% mermó hasta aproximadamente el 10%, cuando el Estado se dio cuenta, con retraso, de que había nacimientos que no habían sido notificados. Según parece, muchas madres no consideraron que tuvieran razón alguna para notificar nuevos nacimientos[23], ya que éstos no incrementaban sus prestaciones sociales.

Este es un ejemplo del problema que supone el modo de recopilar las estadísticas, no supervisamos a todas las personas que pensamos que estamos supervisando. Algunas veces es más difícil identificar ciertos errores de razonamiento, pero podemos

mejorar con la práctica. Vamos a echar un vistazo a una herramienta básica que se emplea mal a menudo.

El gráfico de sectores, o de tarta, es un modo fácil de visualizar porcentajes, cómo se disponen las distintas partes de un todo. Puede que nos interese saber qué porcentaje del presupuesto de educación de una región se gasta en salarios, en materiales de instrucción y en mantenimiento. O puede que nos interese conocer qué porcentaje del dinero que se gasta en materiales de instrucción se emplea en matemáticas, ciencias, lengua, arte, deporte, música, etc. La regla esencial del gráfico de sectores es que los porcentajes tienen que sumar 100. Imaginad una tarta real, si hay nueve personas y cada una quiere un pedazo igual, no podemos cortar la tarta en ocho pedazos, cuando hayáis terminado de cortar la tarta ya no queda nada que cortar. Esto no impidió a Fox News publicar el siguiente gráfico de sectores:



Primera regla de los gráficos de sectores: los porcentajes han de sumar 100 (Fox News, 2012).

Fuente: Opinions Dynamic.

Lo anterior puede suceder, se ofrece a los votantes la opción de apoyar a más de un candidato, pero en ese caso no debiéramos presentar los resultados con un gráfico de sectores.

---

10 Se trata de una cita casi literal de J. Best (2005), «Lies, Calculations and Constructions: Beyond *How to Lie with Statistics*», *Statistical Science,* 20(3), 210-214.

11 Y. Wang (25 de marzo de 2013). Hay más personas con móvil que con W.C., afirma un estudio de la ONU. http://newsfeed.time.com/2013/03/25/more-people-have-cell-phones-than-toilets-u-n-study-shows/.

12 G. Steinem (1992), *Revolution from Within,* Nueva York, Little, Brown. N. Wolf (1991), *The Beauty Myth,* Nueva York, William Morrow.

13 Este ejemplo proviene de J. Best (2005), «Lies, calculations and constructions: beyond *How to Lie with Statistics*», *Statistical Science,* 20(3), 210-214. Las estadísticas están disponibles en www.cdc.gov.

14 Quizá trabajéis en el departamento de pagos de una gran empresa. Un empleado solicita el reembolso de gasolina de su coche de empresa por valor de 5.000 dólares durante el mes de abril. Empecemos con un poco de conocimiento del mundo: hoy en día la mayoría de los automóviles hacen más de 32 kilómetros por galón (3,7 litros) (algunos multiplican ese rendimiento varias veces). También sabemos que lo más rápido que es posible conducir de modo razonable es aproximadamente a 115 km por hora y que si hubiera que conducir durante diez horas al día, todas ellas por autopista, significaría unos 1.125 km por día. Supongamos que mantenemos ese promedio durante el mes laboral de 21,5 días y obtenemos unos 24.220 km. Al hacer estas estimaciones lo habitual es redondear para facilitar los cálculos, así que digamos 24.000. Lo dividimos por el promedio de consumo y obtenemos que el empleado debe haber necesitado unos 750 galones (2.839 litros) de combustible. Buscamos el precio medio nacional del combustible para abril y encontramos que era de 2,89 dólares, por redondear 3 dólares (y concedemos al empleado el beneficio de la duda, puede que no haya conseguido siempre el mejor precio posible cada vez que llenó el depósito). 3 dólares por galón por 750 galones = 2.250 dólares. Ahora los 5.000 dólares de gastos en combustible no resultan ni remotamente verosímiles. Incluso si el empleado hubiera conducido veinticuatro horas al día, el coste no habría sido tan elevado. https//www.fueleconomy.gov/feg/best/bestworstNF.shtml, consultado el 1 de agosto de 2015. http://fuelgaugereport.com/.

15 L. Pollack y H. Weiss (1984), «Communication satellites: countdown for Intelsat VI», *Science,* 223 (4636), 533.

16 Me imagino una historia que podría hacer que esto resultase cierto. A lo mejor había un artilugio que valía 1 dólar y ahora, como parte de una gran campaña de promoción, hay una empresa que no solo está dispuesta a regalártelo, sino a *pagarte* 11.999 dólares para que te lo lleves (eso es una reducción del 12.000%). Estas cosas pasan en los grandes negocios inmobiliarios. A lo mejor es necesario derribar un edificio viejo para poder construir otro; el propietario podría estar pagando un impuesto de la propiedad enorme, el coste de derribar el edificio es alto, así que el propietario está dispuesto a que alguien se lo quite de las manos. En algún momento de finales de los noventa algunas inmobiliarias muy endeudadas estaban vendiendo por 0 dólares, siempre que el nuevo propietario asumiera la deuda.

17 C. Bailey y M. Clarke (2008), «Aligning business leadership development with business needs: the value of discrimination», *Journal of Management Development,* 27(9), 912-934.

Otros ejemplos de reducción del 200%: B. S. Rajashekar y V. P. Kalappa (2006), «Effects of planting seasons on seed yield & quality of tomato varieties resistant to leaf curl virus», *Seed Research,* 34(2), 223-225. http://www.bostoncio.com/AboutRichardCohen.asp.

18 Ilustración © 2016 por Dan Piraro, basado en un ejemplo de Huff, *ibid*.

19 Agradezco al abogado James P. Scanlan, de Washington DC, que respondiera a mi solicitud a la American Statistical Association y me indicara este tipo de error.

20 Este ejemplo tiene su origen en L. Spirer, H. F. Spirers y A. J. Jaffe (1987), *Misused Statistics,* Nueva York, Marcel Dekker, p. 194.

J. Miller (29 de diciembre de 1996), «Los altos costes causan el cierre de una fábrica textil», *New York Times,* Connecticut Section. Y n.d. (12 de enero de 1997), «Corrección», *New York Times,* Connecticut Section.

21 K. J. McLarin (5 de diciembre de 1993), «Toma y daca en las ayudas de Nueva Jersey; se abona a las madres la universidad, pero no hay ayudas para los recién nacidos», *New York Times*.

Véase también: M. Henneberger (11 de abril de 1995), «Repensar las ayudas: desalentar nuevos nacimientos —informe especial; la ayuda del estado es recortada, pero ¿cuáles son las consecuencias?», *New York Times*.

22 *Ibid*.

23 *Ibid*.

DIVERTIRSE CON LOS PROMEDIOS

Un promedio puede ser un útil estadístico de resumen, incluso más fácil de comprender que un gráfico de sectores, permitiéndonos describir una gran cantidad de información con una única cifra. Puede que nos interese saber el patrimonio medio de las personas que están en una sala, para determinar si a nuestros agentes financieros o responsables de ventas podría interesarles reunirse con ellos. O podemos desear conocer el precio medio del combustible para estimar cuánto nos costará conducir desde Vancouver hasta Banff. Pero los porcentajes pueden ser engañosamente complejos.

Hay tres maneras de calcular un promedio y frecuentemente arrojan resultados diferentes, en consecuencia las personas con conocimientos estadísticos habitualmente evitan el término *promedio* y prefieren los términos *media, mediana* y *moda*. No decimos «promedio medio» o «promedio mediano», ni siquiera «porcentaje», decimos *media, mediana* o *moda*. En algunos casos son idénticos, pero en muchos otros no. Si vemos la palabra *promedio* aislada, habitualmente significa la media, pero no podemos estar seguros.

La media es el más común de los tres y se calcula sumando todas las observaciones o informes disponibles y dividiéndolos por la cantidad de observaciones o informes. Por ejemplo, el patrimonio medio de las personas que hay en una sala es simplemente el total del patrimonio, dividido por la cantidad de personas. Si en la sala hay diez personas cuyas posesiones equivalen a 100.000 dólares por cada uno, en la sala hay un patrimonio total de 1 millón de dólares, y podemos calcular la media sin necesidad de una calculadora: 100.000 dólares. Si en otra sala hay diez personas cuyo patrimonio neto varía entre 50.000 y 150.000 dólares, pero el total continúa siendo 100.000 dólares, la media aún es 100.000 dólares (ya que nos limitamos a

tomar el total de 1 millón de dólares, y lo dividimos entre diez personas, independientemente de lo que tenga cada una de ellas).

La mediana es el número del medio en un conjunto de números (los estadísticos llaman a ese conjunto una «distribución»): la mitad de las observaciones se sitúan por encima y la mitad por debajo. Recordemos que el objetivo de un promedio es permitir representar un conjunto de datos mediante una única cifra. La mediana nos permite alcanzar mejor ese objetivo cuando algunas de nuestras observaciones son muy diferentes de la mayoría, lo que los estadísticos llaman *valores atípicos*.

Si entramos en una sala en la que hay nueve personas e imaginamos que ocho de ellas poseen un patrimonio neto próximo a los 100.000 dólares y una persona está en bancarrota, con un patrimonio neto negativo de 500.000 dólares a causa de sus deudas, la siguiente sería la composición de la sala:

Persona 1: –500.000 dólares
Persona 2: 96.000 dólares
Persona 3: 97.000 dólares
Persona 4: 99.000 dólares
Persona 5: 100.000 dólares
Persona 6: 101.000 dólares
Persona 7: 101.000 dólares
Persona 8: 101.000 dólares
Persona 9: 104.000 dólares

La suma total asciende a 299.000 dólares, la dividimos por el total de las observaciones, nueve, y la media resultante es de 33.222 dólares por persona. Pero la media no nos permite describir bien el conjunto de la sala. Sugiere que nuestro gestor de inversiones no estaría interesado en visitar a esas personas, cuando lo que ocurre en realidad es que hay una única persona extraña, un valor atípico que disminuye el promedio. Este es el problema de la media: es sensible a los valores atípicos.

En este caso, la mediana sería 100.000 dólares: cuatro personas ganan menos de esa cantidad y otras cuatro ganan más. La moda es 101.000 dólares, la cifra que aparece con más frecuencia. En este ejemplo concreto, la mediana y la moda son más ilustrativas.

Hay muchas maneras de utilizar los promedios para manipular lo que deseamos que los demás vean en nuestros datos.

Supongamos que dos amigos y tú habéis fundado una empresa con cinco empleados. Es el final del año y queréis informar a los empleados de la situación financiera de la empresa para que puedan reconciliarse con las largas horas de trabajo y las pizzas frías que se han comido, y para atraer inversores. Pongamos que cuatro empleados, los programadores, han ganado cada uno de ellos 70.000 dólares, y un empleado, el recepcionista/encargado de mantenimiento, ha ganado 50.000 dólares. Eso arrojaría un salario (medio) por empleado de 66.000 dólares al año (4 × 70.000) + (1 × 50.000), divido entre 5. Tú y tus dos amigos ganasteis 100.000 al año de sueldo. En consecuencia, vuestra nómina asciende a (4 × 70.000) + (1 × 50.000) + (3 × 100.000) = 630.000 dólares. Pongamos que vuestra empresa obtuvo 210.000 dólares de beneficios y que los dividís a partes iguales entre los emprendedores fundadores como bonificación, lo que os proporciona 100.000 + 70.000 dólares a cada uno de vosotros. ¿Cómo informar sobre estos datos?

Podríais decir:

Salario promedio de los empleados: 66.000 dólares

Salario promedio + bonificación de los propietarios: 170.000 dólares

Lo anterior es cierto, pero probablemente no tiene muy buena pinta, salvo para vosotros y vuestras mamás. Si vuestros empleados se enteran, podrían sentirse infravalorados; los inversores potenciales podrían pensar que los socios fundadores están sobrecompensados. Por lo tanto, podríais informar de lo siguiente:

Salario promedio de los empleados: 66.000 dólares
Salario promedio de los propietarios: 100.000 dólares
Beneficios: 210.000 dólares

Esto ya tiene mejor pinta para los inversores potenciales. Además podríais eliminar el dato de que repartisteis los beneficios entre los propietarios y también eliminar la última línea, la parte sobre los beneficios, cuando informéis a los empleados. Los cuatro programadores van a pensar que son muy valorados, ya que ganan más del promedio. Vuestro recepcionista no estará tan contento, pero seguramente ya sabía que los programadores ganan más que él.

Ahora vamos a imaginar que tienes la sensación de estar trabajando demasiado y quieres persuadir a tus dos socios, que no saben mucho sobre pensamiento crítico, de que es necesario contratar más empleados. Podrías, igual que muchas otras empresas, informar de los «beneficios por empleado», dividiendo el beneficio de 210.000 dólares entre los cinco empleados:

Salario promedio de los empleados: 66.000 dólares
Salario promedio de los propietarios: 100.000 dólares
Beneficio promedio por empleado: 42.000 dólares

Ahora podrías afirmar que el 64% de los salarios que pagáis a los empleados (42.000/66.000) retorna a vosotros en forma de beneficios, lo que significa que solo tenéis que pagar el 36% de los salarios tras agregar los beneficios. Es evidente que no hay nada en estos datos que lleve a pensar que incorporar un empleado nuevo va a aumentar los beneficios, vuestros beneficios pueden no depender de cuántos empleados haya, pero para alguien que no piensa críticamente, esto suena como una razón sólida para contratar más empleados.

Por último, ¿y si queremos afirmar que sois unos patronos inusitadamente justos y equitativos y la diferencia entre lo que ganáis como beneficios y lo que ganan vuestros empleados es bastante sensata? Tomáis los 210.000 dólares de beneficios y distribuís 150.000 dólares de los mismos como bonus salarial para vosotros y vuestros socios, dejando los 60.000 dólares

restantes como «beneficios». Ahora calculáis el salario promedio, pero os incluís a vosotros y vuestros socios con los bonus salariales.

Salario promedio: 97.000 dólares
Beneficio promedio de los propietarios: 20.000 dólares

Y ahora vamos a divertirnos de verdad:

Total de costes salariales incluyendo los bonus: 840.000 dólares
Salarios: 780.000 dólares
Beneficios: 60.000 dólares

Ahora tiene una pinta mucho mejor, ¿no os parece? Del total de 840.000 dólares disponible para salarios y bonus, solo 60.000, el 7%, ha ido a los beneficios de los propietarios. Vuestros empleados van a pensar que sois irreprochables: ¿quién estaría resentido con un empresario que se limita a quedarse con un 7%? Y, de hecho, ni siquiera asciende a eso, el 7% se divide entre los propietarios de la empresa, arrojando un 2,3% para cada uno. ¡No hay queja posible!

Se puede hacer aún mejor. Imaginad que en el primer año de vuestra empresa únicamente tuvisteis empleados a tiempo parcial que ganaban 40.000 dólares anuales. Al año siguiente únicamente tuvisteis empleados a tiempo completo que ganaban los 66.000 dólares que mencionamos antes. Podéis afirmar con toda honestidad que el salario promedio de los empleados subió un 65%. ¡Qué magníficos empresarios! Lo malo es que en este caso estaríais agregando trabajo a tiempo parcial y a tiempo completo, pero no seríais los primeros: U.S. Steel ya lo hizo allá por los años cuarenta.

En las causas penales, la forma en que se presenta la información, la acusación, tiene un profundo efecto sobre las conclusiones del jurado en torno al veredicto. Aunque matemáticamente sean equivalentes[24], es mucho más persuasivo testificar que «la probabilidad de que la sangre del sospechoso no sea la misma que la de las gotas de sangre es de

0,1%» (una entre mil), que decir «una de cada mil personas en Houston también tiene el mismo tipo de sangre».

Los promedios se emplean a menudo para expresar resultados, como «uno de cada X matrimonios acaba en divorcio». Pero eso no significa que la estadística se aplique a nuestra calle, a nuestro club de bridge o a cualquiera que conozcamos. Puede que sí o puede que no, se trata de un promedio nacional y puede haber ciertos *factores de vulnerabilidad* que contribuyan a pronosticar quién se divorciará y quién no.

En el mismo sentido, puede que leamos que uno de cada cinco niños que nacen es chino. Observáis que la sueca del final de la calle ya tiene cuatro niños y que la madre espera otro bebé. Pero eso no significa que vaya a dar a luz un niño chino —el de los cinco del promedio de todos los nacimientos mundiales—, ni tampoco que eso vaya a producirse entre los nacimientos de un cierto edificio o de un barrio o siquiera de un país.

Hay que tener cuidado con los promedios y cómo se aplican. Un modo de ser engañados es cuando el promedio combina muestras de poblaciones dispares. Esto puede dar lugar a observaciones como la siguiente:

> En promedio los humanos tienen un testículo[25].

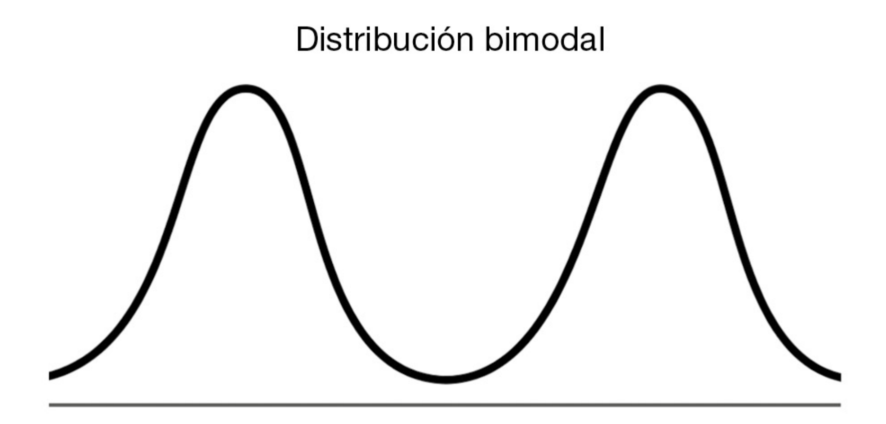
Este ejemplo ilustra la diferencia entre media, mediana y moda. Dado que en el mundo hay algunas mujeres más que hombres, la mediana y la moda son cero, en tanto que la media se sitúa próxima a uno (0,98, más o menos).

También es necesario recordar que el promedio no nos dice nada sobre el rango. La temperatura promedio anual en el Valle de la Muerte, en California, son unos agradables 25 grados C. Pero el rango puede matarnos, con temperaturas registradas que oscilan entre –9,4 y 56,6 grados[26].

O... podría deciros que la riqueza neta promedio de cien personas en una sala son unos alucinantes 350 millones de dólares. Seguramente pensaríais que es justo el lugar para soltar a cien de vuestros mejores vendedores. Pero en la sala podrían

estar Mark Zuckerberg (riqueza neta: 35.000 millones de dólares) y noventa y nueve indigentes. El promedio puede borrar diferencias que son importantes.

Otro asunto que hay que vigilar en los promedios es la *distribución bimodal*. Recordad que la *moda* es el valor que ocurre con mayor frecuencia. En muchos conjuntos de datos biológicos, físicos y sociales, la distribución tiene dos o más picos, es decir, dos o más valores que aparecen con más frecuencia que los demás.



Por ejemplo, un gráfico como este podría mostrar la cantidad de dinero gastado en almuerzos durante una semana[27] (eje *x*) y cuántas personas gastaron cada cantidad (eje *y*). Imaginad que en vuestra investigación hay dos grupos de personas diferentes, niños (joroba de la izquierda, compran almuerzos escolares) y ejecutivos de empresas (joroba de la derecha, van a restaurantes de moda). En este caso, la media y la mediana podrían ser una cifra situada justo entre los dos y no nos informarían sobre lo que realmente está sucediendo; de hecho, la media y la mediana a menudo son cantidades que nadie está usando. Un gráfico como este con frecuencia nos proporciona un indicio de que nuestra muestra es heterogénea, o que estamos comparando naranjas con manzanas. Lo mejor en este caso es informar de que hay una distribución bimodal y dar cuenta de las dos modas. Mejor aún, subdividid el grupo en dos y ofreced las estadísticas de cada grupo resultante.

Hay que ser cuidadosos al extraer conclusiones sobre los individuos y los grupos basándonos en promedios. Las trampas

son tan frecuentes que tienen nombre: la falacia ecológica y la falacia de la excepción. La falacia ecológica tiene lugar cuando hacemos inferencias sobre un individuo basándonos en datos agregados (como la media de un grupo) y la falacia de la excepción sucede cuando hacemos inferencias sobre un grupo basándonos en datos sobre unos pocos individuos excepcionales.

Por ejemplo, imaginemos dos pueblos pequeños, cada uno de cien habitantes. El pueblo A tiene noventa y nueve personas que ganan 80.000 dólares al año y una persona superrica que encontró petróleo en su finca y que gana 5.000.000 dólares al año. El pueblo B tiene cincuenta personas que ganan 100.000 dólares al año y cincuenta personas que ganan 140.000. El ingreso medio del pueblo A es de 129.000 dólares y el del pueblo B de 120.000. Aunque el pueblo A tiene un ingreso medio superior, el 99% de las veces que seleccionemos al azar a un individuo del pueblo B tendrá unos ingresos superiores a los de un individuo seleccionado al azar del pueblo A. La falacia ecológica consiste en creer que si hemos seleccionado al azar un individuo del grupo con la media más alta, ese individuo debería tener unos ingresos superiores. Lo interesante es que en los ejemplos anteriores, aunque la *media* es superior en el pueblo A, la *moda* es superior en el pueblo B. (Esto no siempre es así).

Otro ejemplo, se ha sugerido que las personas ricas tienen mayores probabilidades de votar a los republicanos, pero los datos indican que los estados más ricos tienden a votar a los demócratas. La riqueza de los estados más ricos podría estar sesgada por un pequeño porcentaje de individuos superricos. Durante la elección presidencial de 2004[28], el candidato republicano, George W. Bush, ganó en los quince estados más pobres, y el demócrata, John Kerry, ganó en nueve de los once estados más ricos. Sin embargo, el 62% de quienes obtenían ingresos anuales por encima de los 200.000 dólares votaron por Bush, mientras que tan solo el 36% de los votantes con ingresos anuales de 15.000 dólares o menos votaron por Bush.

Como ejemplo de la falacia de la excepción, puede que hayáis leído que los Volvo están entre los automóviles más fiables y decidáis comprar uno. De camino al concesionario, pasáis frente

a un taller mecánico de Volvo y veis un aparcamiento completo lleno de Volvo averiados. Si cambiáis de opinión sobre la compra basándoos en esto, estaréis usando una cantidad relativamente pequeña de casos excepcionales para realizar una inferencia sobre el grupo entero. Nadie ha afirmado que los Volvo nunca se averíen, tan sólo que en conjunto tienen menos averías. (De aquí la omnipresente advertencia de que «el rendimiento de cada uno puede variar»). Hay que señalar también que lo que habéis visto os está influyendo exageradamente de otro modo: el lugar donde más frecuentemente se verán Volvo averiados es un taller mecánico de Volvo. Vuestra «tasa de base» se ha desplazado y ya no podéis considerar que se trate de una muestra al azar.

Ahora que ya sois expertos en promedios, no deberíais caer en la conocida confusión de creer que las personas tendían a vivir menos tiempo hace cien años. Lo más probable es que hayáis leído que la esperanza de vida se ha incrementado gradualmente en los tiempos modernos. Para quienes nacieron en 1850, la esperanza de vida promedio era de treinta y ocho años para los hombres y cuarenta para las mujeres[29], y para quienes nacieron en 1990, es setenta y dos y setenta y nueve años respectivamente. En consecuencia, hay una tendencia a creer que en el siglo XIX no había tanta gente de cincuenta y sesenta años, ya que las personas no vivían tanto tiempo. Sin embargo, la realidad es que sí vivían tanto tiempo, lo que sucede es que la mortalidad en los recién nacidos y en la infancia era tan elevada que sesgaba el promedio. Si uno llegaba a los veinte años, entonces ya era posible vivir mucho tiempo. De hecho, en 1850 una mujer blanca de cincuenta años podía esperar vivir hasta los setenta y tres y medio y una de sesenta hasta los setenta y siete. Sin duda, la esperanza de vida se ha elevado en unos diez años para quienes hoy tienen cincuenta o sesenta años, en comparación con 1850, en gran medida gracias a los mejores servicios de salud. Pero igual que sucedía con los anteriores ejemplos sobre la sala llena de personas con ingresos muy diferentes, los cambios en el promedio de esperanza de vida al nacer durante los últimos 175 años reflejan diferencias

significativas en las dos muestras: había muchas más muertes en la infancia en aquellos tiempos y eso rebajaba el promedio.

Un rompecabezas: el niño promedio usualmente no viene de la familia promedio[30]. ¿Por qué? A causa del desplazamiento de las líneas de base. (En este comentario estoy utilizando «promedio» en lugar de «media» gracias a un magnífico artículo sobre el tema de James Jenkins y Terrell Tuten, quienes lo emplearon en su título).

Ahora vamos a suponer que hemos leído que el promedio de niños por familia en una comunidad suburbana es de tres. En consecuencia, podríais concluir que un niño promedio tendrá dos hermanos. Pero sería un error. El mismo problema lógico tiene lugar si nos preguntamos si el alumno universitario promedio asiste a una universidad de tamaño promedio, si el empleado promedio gana el salario promedio o si el árbol promedio está en el bosque promedio. ¿Y eso?

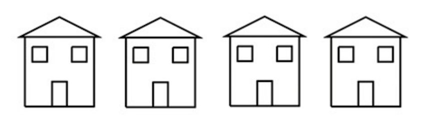
Todos estos casos implican un desplazamiento de la línea de base o grupo de muestreo que estudiamos. Cuando calculamos la cantidad promedio de niños por familia estamos muestreando familias. Tanto una familia muy grande como una pequeña cuentan como una familia. Cuando calculamos la cifra promedio (media) de hermanos, estamos muestreando niños. Cada niño de una familia grande se cuenta una vez, por lo tanto la cantidad de hermanos que tiene cada uno de ellos tiene un peso importante en el promedio de cantidad de hermanos. En otras palabras, una familia con diez niños cuenta una sola vez en la estadística de promedio de *familia,* pero cuenta diez veces en la estadística de promedio de *cantidad de hermanos*.

Supongamos que en un barrio de esa comunidad hipotética hay treinta familias. Cuatro familias no tienen hijos, seis familias tienen un hijo, nueve familias tienen dos hijos y once familias tienen seis hijos. La cantidad promedio de hijos por familia es de tres, ya que noventa (la cantidad total de hijos) se divide entre treinta (la cantidad total de familias). Echemos un vistazo a la cantidad promedio de hermanos[31]. El error que la gente suele cometer es pensar que si la familia promedio tiene tres hijos, entonces cada niño ha de tener dos hermanos de promedio. Pero

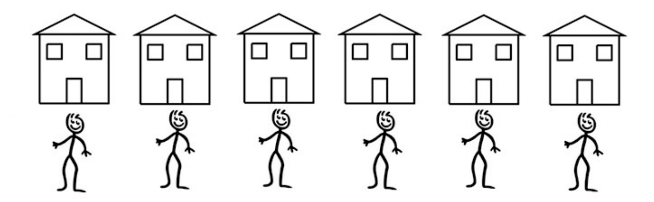
en las familias de hijo único, cada uno de esos seis niños tiene cero hermanos. En las familias de dos hijos, cada uno de los dieciocho niños tiene un hermano. En las familias de seis hijos, cada uno de los sesenta y seis niños tiene cinco hermanos. Entre los 90 niños hay 348 hermanos. En consecuencia, aunque el *niño* promedio viene de una familia de tres hijos, hay 348 hermanos divididos entre 90 niños, un promedio de casi cuatro hermanos por niño.

	Familias	*Hijos/Familia*	*Total de hijos*	*Hermanos*
	4	0	0	0
	6	1	6	0
	9	2	18	18
	11	6	66	330
Totales	30		90	348

Promedio de hijos por familia: 3,0



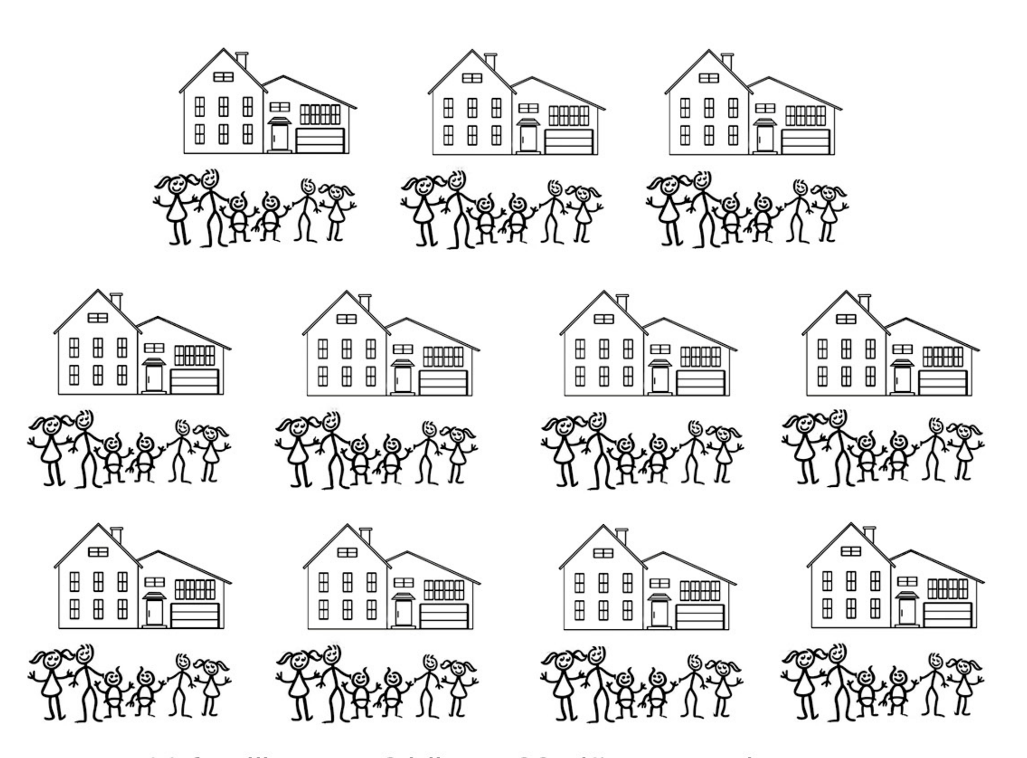
4 familias con 0 hijos



6 familias con 1 hijo – 6 niños con 0 hermanos



9 familias con 2 hijos – 18 niños con1 hermano



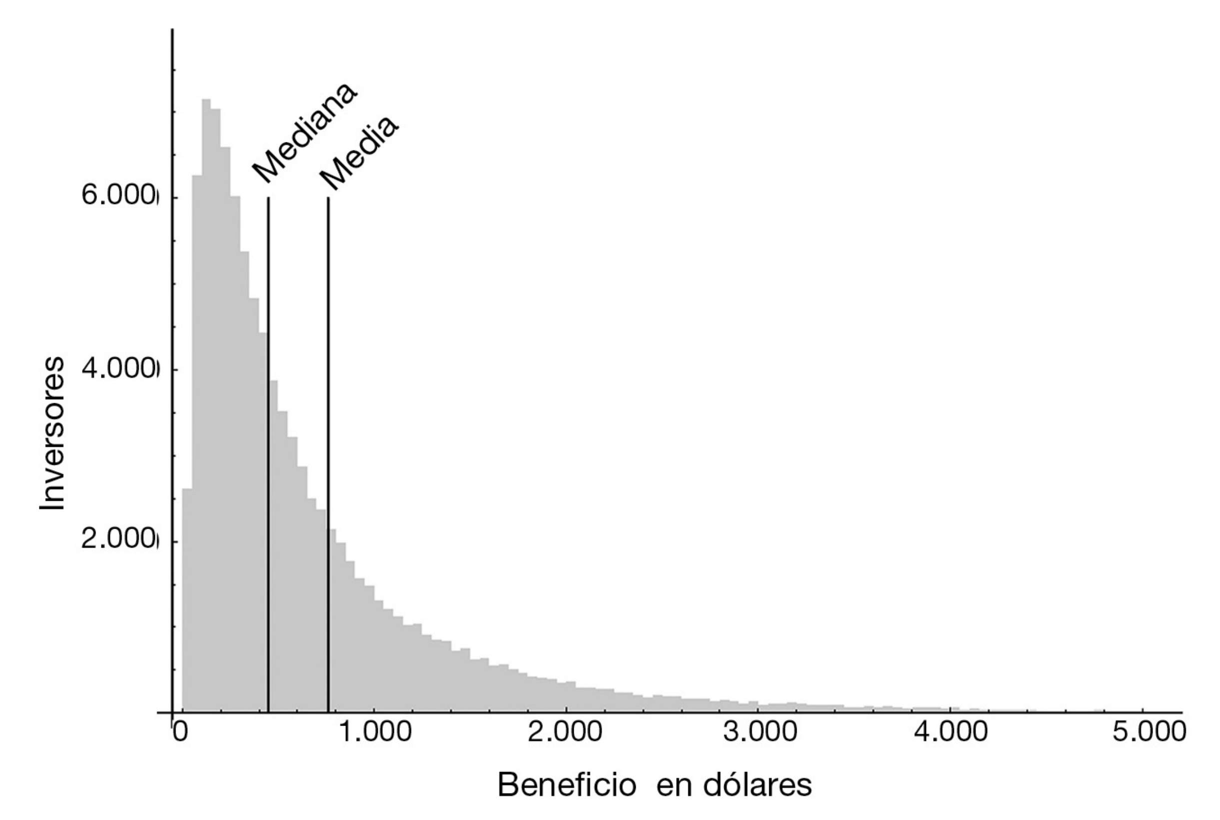
11 familias con 6 hijos – 66 niños con 5 hermanos

Promedio de hermanos por niño: 3,9

Veamos ahora el tamaño de la universidad. En los Estados Unidos hay muchas universidades muy grandes (como Ohio State

y Arizona State) con más de 50.000 estudiantes. También hay muchas universidades pequeñas, con menos de 3.000 estudiantes (como Kenyon College y Williams College). Si contamos las *universidades,* posiblemente encontremos que la universidad promedio tiene 10.000 estudiantes. Pero si contamos el alumnado, encontraremos que el estudiante promedio va a una universidad que tiene más de 30.000 alumnos. Esto se debe a que al contar los estudiantes, obtenemos muchos más datos de las universidades grandes. De manera semejante, la persona promedio no vive en una ciudad promedio y el golfista promedio no consigue la ronda promedio del campo (el total de golpes para los dieciocho hoyos).

Los ejemplos anteriores implican un desplazamiento de la línea de base o denominador. Vamos a considerar otro sobre el tipo de distribución sesgada que observamos antes en relación a la mortalidad infantil: el inversor promedio no obtiene el beneficio promedio[32]. En un estudio, el beneficio promedio de una inversión de 100 dólares que se mantenga durante treinta años fue de 760 dólares o un 7% anual. Pero el 9% de los inversores perdieron dinero y un llamativo 69% no llegó a ganar el beneficio promedio. Esto se debe a que el promedio está sesgado hacia unas pocas personas que obtuvieron cifras muy por encima del promedio, en la figura siguiente la *media* se desplaza hacia la derecha gracias a esos inversores prósperos que ganaron fortunas.



Beneficios de una inversión de 100 dólares durante treinta años. Obsérvese que la mayoría de las personas obtienen menos de la media de beneficio y unos pocos afortunados obtienen más de cinco veces el beneficio medio.

---

24 J. J. Koehler (2001), «The psychology of numbers in the courtroom: how to make DNA-match statistics seem impressive or insufficient», *Southern California Law Review,* 74, 1275-1305. Y J. J. Koehler (2001), «When people are persuaded by DNA match statistics?», *Law and Human Behavior,* 25(5), 493-513.

25 Atribuida al profesor de matemáticas Desmond MacHale del University College, Cork, Irlanda.

26 http://en.wikipedia.org/wiki/Death_Valley.

27 A modo de ejemplo supongamos que seis adultos gastan las siguientes cantidades en sus almuerzos (12, 10, 10, 12, 11, 11 dólares) y seis niños gastan (4, 3,85, 4,15, 3,50, 4,50, 4 dólares). La mediana (para un número par de observaciones a veces la mediana se considera la media entre las dos cifras centrales o, en este caso, la media entre 4,5 y 10) es 7,25 dólares. La media y la mediana son cifras que realmente nadie ha gastado.

28 Véase A. Gelman (2008), *Red State, Blue State, Rich State, Poor State,* Princeton, NJ, Princeton University Press.

29 Estas cifras corresponden a mujeres y hombres blancos. Los datos para los no blancos no están disponibles. http://www.infoplease.com/ipa/A0005140.html. Una fuente adicional de preocupación es que las cifras de 1850 para los Esados Unidos son exclusivamente del estado de Massachusetts, según la Oficina del Censo.

30 El título de esta sección y los comentarios siguen muy de cerca el trabajo de Jenkins y Tuten: J. Jenkins y J. Tuten (1992), «Why isn't the average child from the average family? And similar puzzles», *American Journal of Psychology,* 105(4), 517-526.

31 La ilustración de los monigotes viene de Etsy, https://www.etsy.com/listing/221530596/stick-figure-family-car-van-bike-funny; casa grande y pequeña dibujadas por el autor; casa mediana de http://www.clipartbest.com/clipart-9TRgq8pac.

32 Una simulación, véase A. Tabarrok (11 de julio de 1924), «Average stock market returns aren't average», http://marginalrevolution.com/marginalrevolution/2014/07average-stock-market-returns-arent-average.html, consultado el 14 de octubre de 2014.

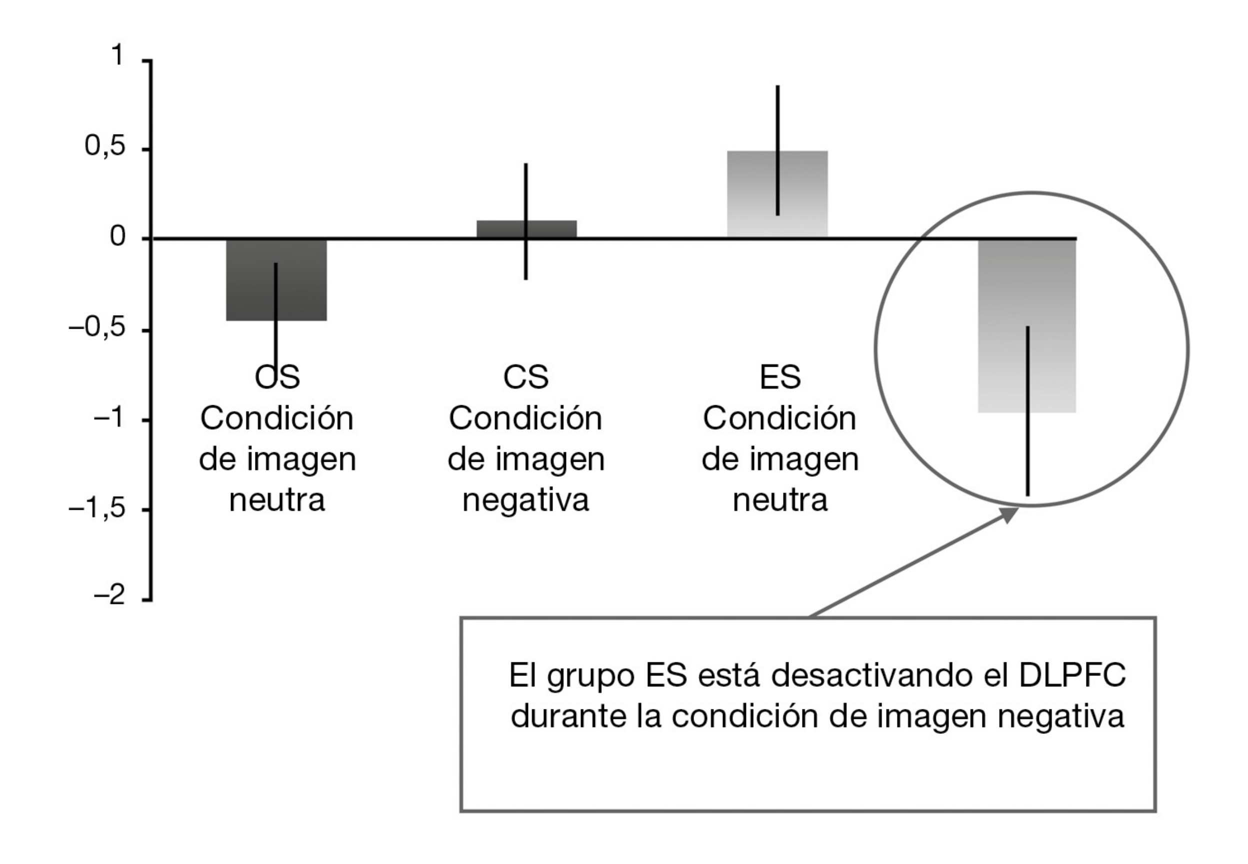
LAS TRAVESURAS DE LOS EJES

El cerebro humano no ha evolucionado para procesar grandes cantidades de datos numéricos presentados en forma de texto; por el contrario, nuestros ojos buscan patrones visuales en los datos. La forma más precisa, pero menos interpretable, de presentación de datos es elaborar una tabla que muestre cada uno de los valores. Sin embargo, a la mayoría de las personas les resulta difícil, si no imposible, detectar patrones y tendencias en tales datos, en consecuencia nos apoyamos en los diagramas y los gráficos. Básicamente hay dos tipos de gráficos: o bien representan visualmente cada dato (como en un diagrama de dispersión), o bien llevan a cabo alguna forma de reducción de los datos, mediante las que los resumimos y ofrecemos, por ejemplo, solamente las medias o las medianas.

Existen muchas maneras de utilizar los gráficos para manipular, distorsionar y tergiversar los datos. El consumidor cuidadoso de información debe evitar ser seducido por ellas.

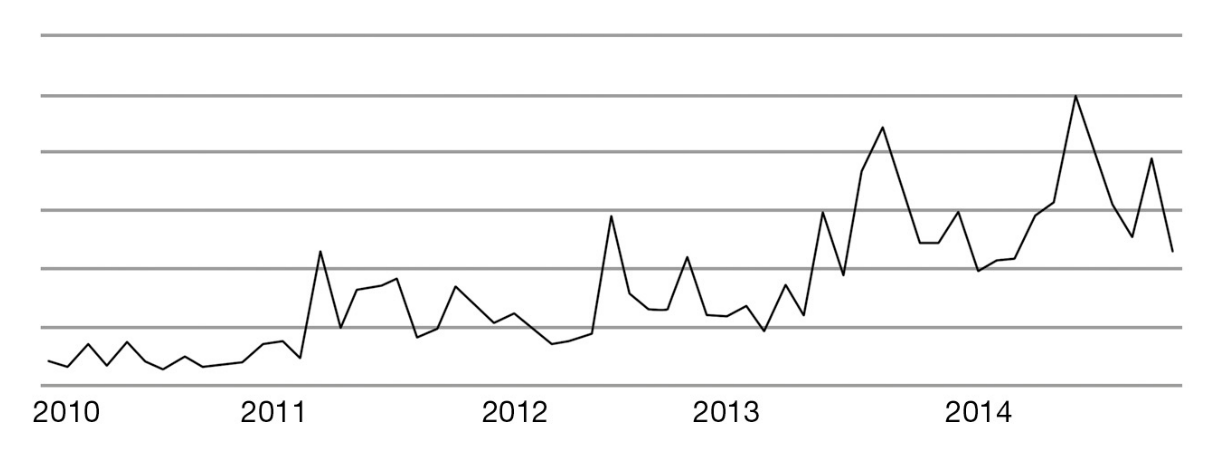
Ejes sin designación

La manera más común de mentir con un gráfico estadístico es no designar los ejes. ¡Si los ejes carecen de designación, es posible representar o ilustrar cualquier cosa! El siguiente es un ejemplo de un póster presentado en un congreso por un investigador novato[33] y aparecía así (lo he redibujado):



¿Qué significa todo esto? A partir del texto del propio póster (que no aparece en este gráfico), sabemos que los investigadores estudian la activación cerebral en pacientes con esquizofrenia (ES). ¿Quiénes son los CS? No nos lo dicen, pero a partir del contexto, están siendo comparados con los ES, podemos suponer que se trata de «sujetos de control saludables». Pues bien, parece haber diferencias entre los CS y los ES, pero, hmmm... el eje *y* tiene cifras, pero... ¡esas unidades podrían ser cualquier cosa! ¿Qué es lo que estamos viendo? ¿Las puntuaciones de una prueba, los niveles de activación cerebral, la cantidad de regiones cerebrales activadas? ¿La cantidad de flanes de la marca Jell-O que han comido, de películas de Johnny Depp que han visto en las últimas seis sema nas? (Para ser honestos he de reconocer que posteriormente los autores publicaron sus resultados en una revista con revisión de expertos y corrigieron su error, después de que una web les indicara el despiste).

En el siguiente ejemplo se trazan las ventas brutas de una editorial[34], eliminando los datos originados por las campañas Kickstarter[35].



Igual que sucedía en el ejemplo anterior, pero en este caso con el eje *x,* tenemos cifras, pero no se nos dice qué representan. En este caso probablemente resulta evidente: podemos suponer que 2010, 2011, etc. se refieren a años o ejercicios fiscales y el hecho de que aparezcan sierras entre los años sugiere que se hace un seguimiento mensual (pero sin las denominaciones adecuadas únicamente podemos suponerlo). El eje *y* está ausente, de modo que no sabemos qué es lo que se mide (¿unidades vendidas o dólares?) y no sabemos qué es lo que representa cada línea horizontal. El gráfico podría representar un incremento de ventas de 5 céntimos al año a 5 dólares al año, o de 50 millones de unidades a 500 millones de unidades. Pero no nos preocupemos, un texto esclarecedor acompaña al gráfico: «Ha sido otro gran año». Vamos a tener que creerles.

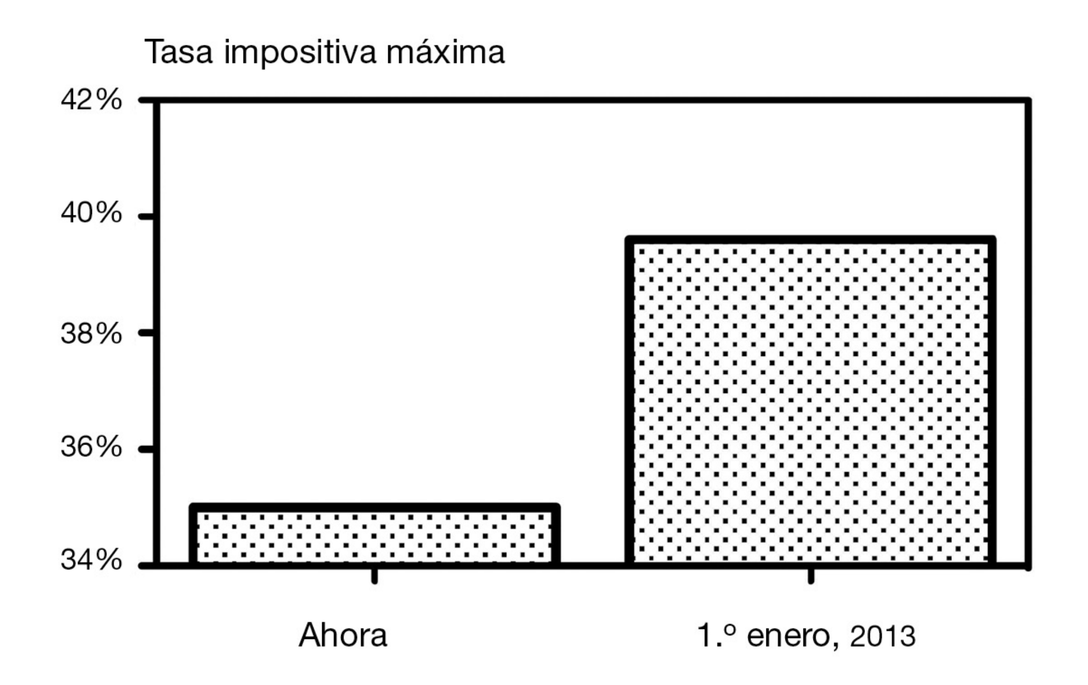
Eje vertical truncado

Un gráfico bien diseñado muestra con claridad los puntos finales de un continuo. Esto resulta especialmente importante cuando pretendemos documentar un cambio real o proyectado en una cantidad y cuando deseamos que los lectores extraigan las conclusiones correctas. Si estamos representando la tasa de crímenes, muertes, nacimientos, ingresos o cualquier cantidad que pudiera tener el valor 0, entonces 0 debiera ser el punto mínimo del gráfico. Pero si lo que pretendemos es crear pánico o escándalo, podemos iniciar nuestro eje *y* en algún punto próximo

al valor más bajo que estemos trazando, esto resaltará la diferencia que deseamos subrayar, ya que arrastramos la mirada hacia el tamaño de la diferencia, tal como se muestra en el gráfico, y se oculta el tamaño real de la diferencia.

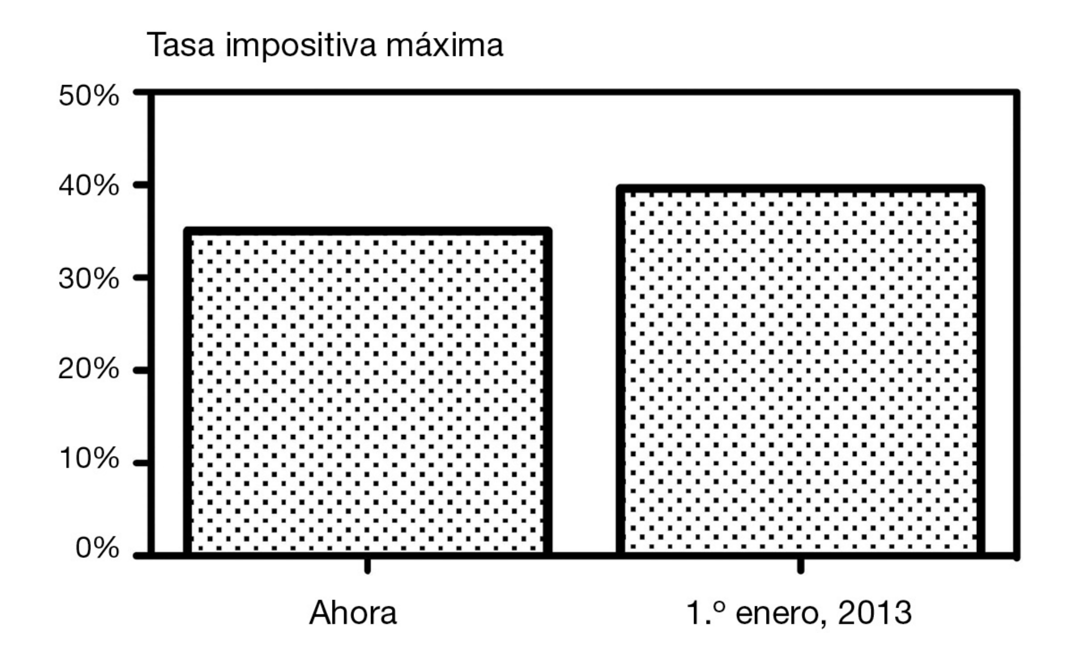
En 2012 el noticiero de Fox divulgó el siguiente gráfico[36] para mostrar qué sucedería si los recortes de impuestos de Bush se terminaban:





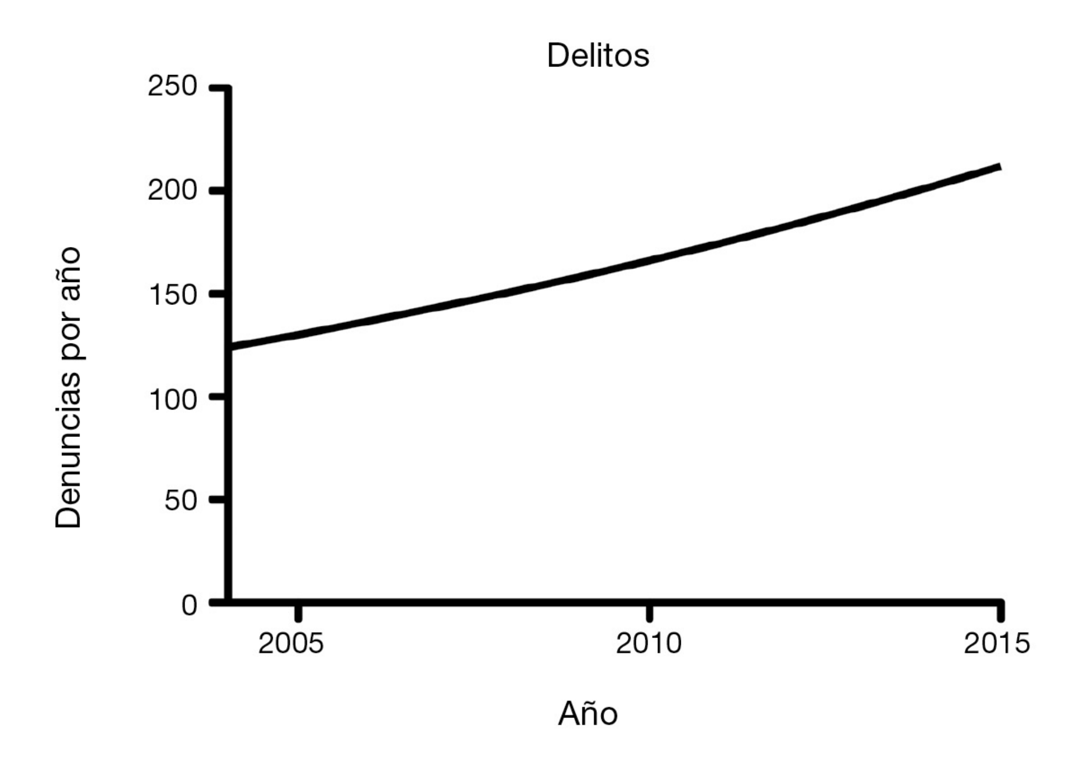
El gráfico genera la impresión visual de que los impuestos subirían mucho: la columna de la derecha tiene un tamaño seis veces mayor que la columna de la izquierda. ¿A quién le apetece que le aumenten los impuestos seis veces? Los televidentes que fueran numerofóbicos pueden no detenerse a examinar el eje para comprobar que la diferencia real es entre una tasa impositiva del 35% y una del 39,6. Es decir, si los recortes terminan los impuestos subirían un 13%, no el factor de seis que se ha representado (un incremento de 4,6 puntos porcentuales es el 13% de 35%).

Si el eje *y* comenzara en cero, el 13% sería visualmente evidente:

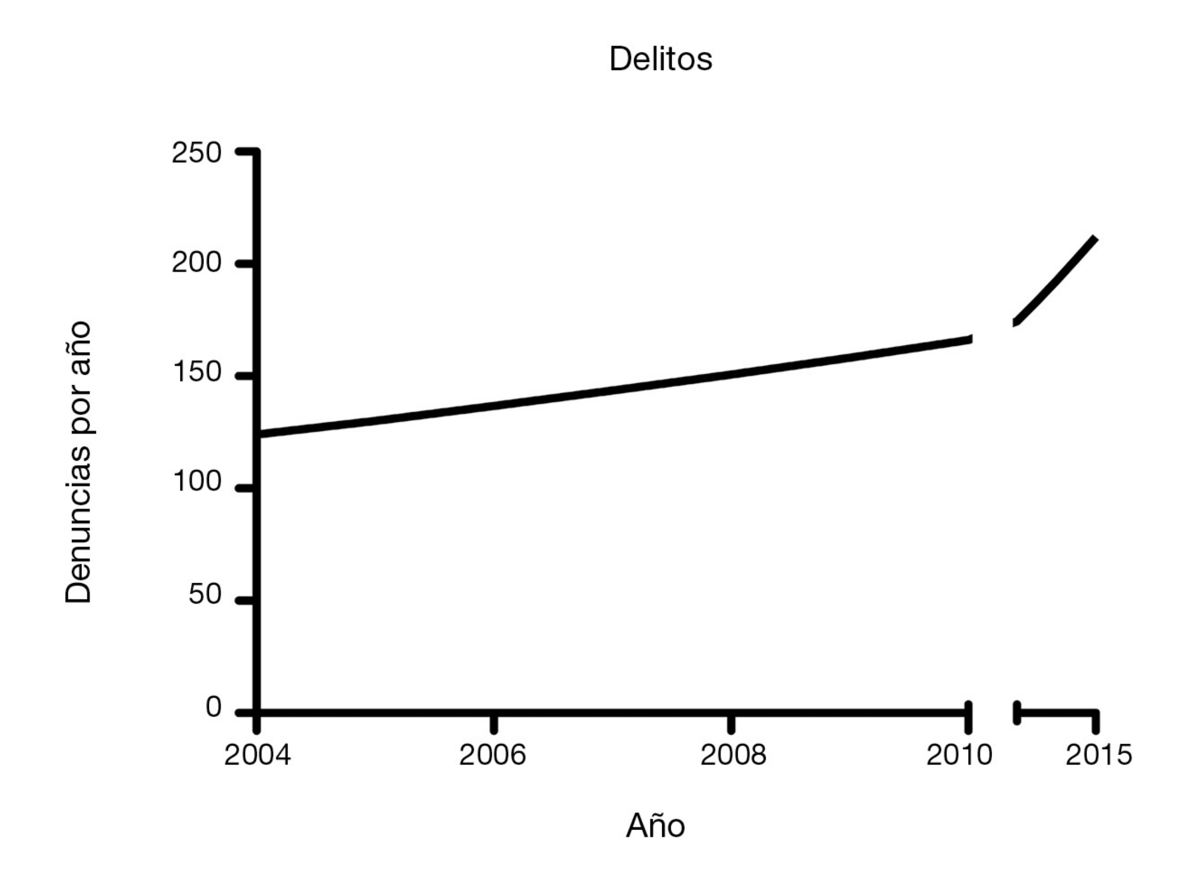


Discontinuidad en un eje vertical o en uno horizontal[37]

Imaginemos una ciudad en la que los delitos han estado aumentando a una tasa del 5% anual durante los últimos diez años. Podríamos representar el gráfico así:

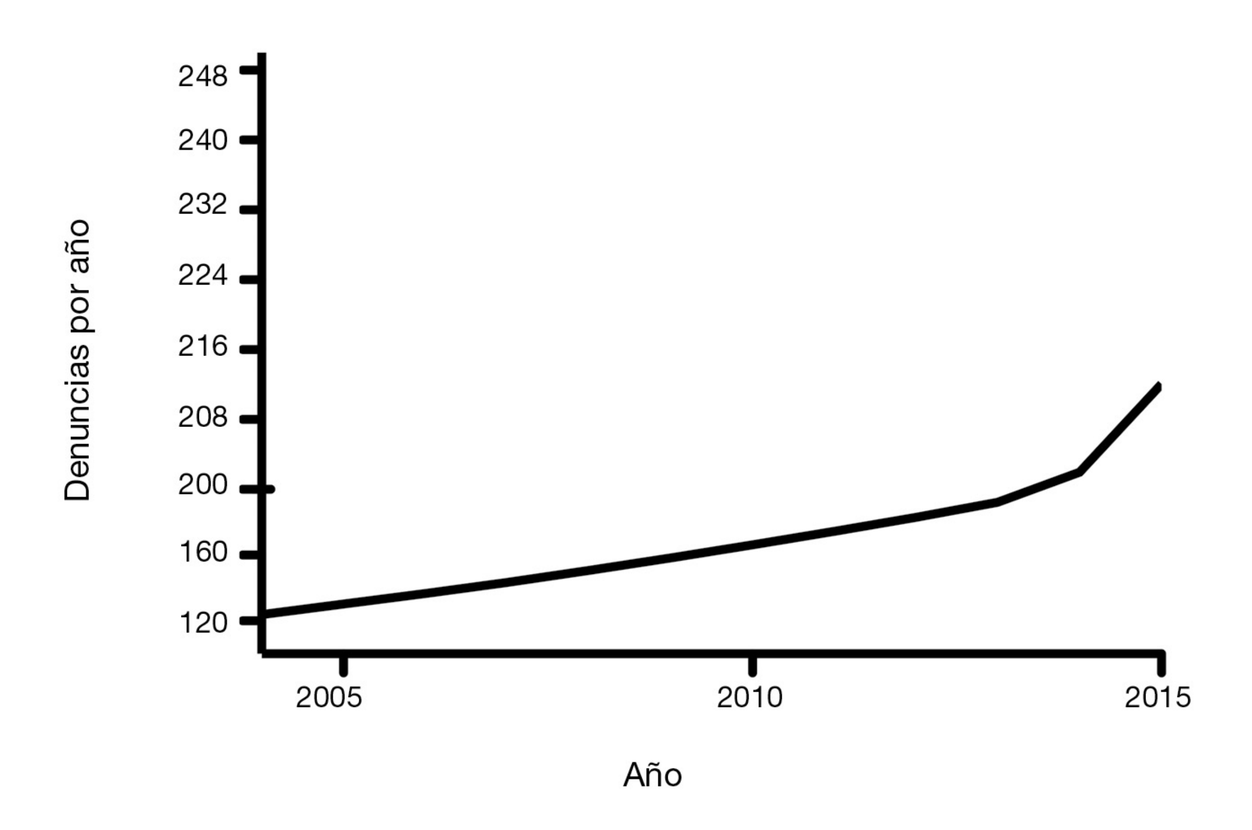


No hay nada de malo; pero supongamos que nuestra meta es vender sistemas de seguridad privada para el hogar y queremos asustar a la gente para que compre nuestro producto. Podríamos utilizar los mismos datos, pero creando una discontinuidad en el eje *x.* Esto supondría distorsionar la verdad y engañaría maravillosamente a la vista:



En este caso la imagen provoca la impresión de que los delitos se han incrementado dramáticamente. Pero sabemos qué es lo que ha sucedido; la discontinuidad en el eje *x* acumula cinco años de datos en la misma extensión de gráfico que se emplea para dos años. No es sorprendente que muestre un incremento evidente. Se trata de un error fundamental en el diseño del gráfico, pero como la mayoría de los lectores no se preocupan de observar con atención los ejes, es fácil que eso pase desapercibido.

No necesitamos limitar nuestra creatividad a la discontinuidad del eje *x;* también podemos obtener el mismo efecto creando una discontinuidad en el eje *y,* después lo ocultamos sin mostrar una ruptura en la línea. Ya que estamos, vamos a truncar el eje *y:*



Esto es algo perverso. La mayoría de los lectores se limitan a observar la curva del gráfico y no perciben que los indicadores de nivel del eje vertical comienzan con intervalos de cuarenta denuncias entre sí, pero repentinamente, a partir del nivel doscientos, indican únicamente ocho denuncias entre cada uno de ellos. ¿Empezamos a divertirnos?

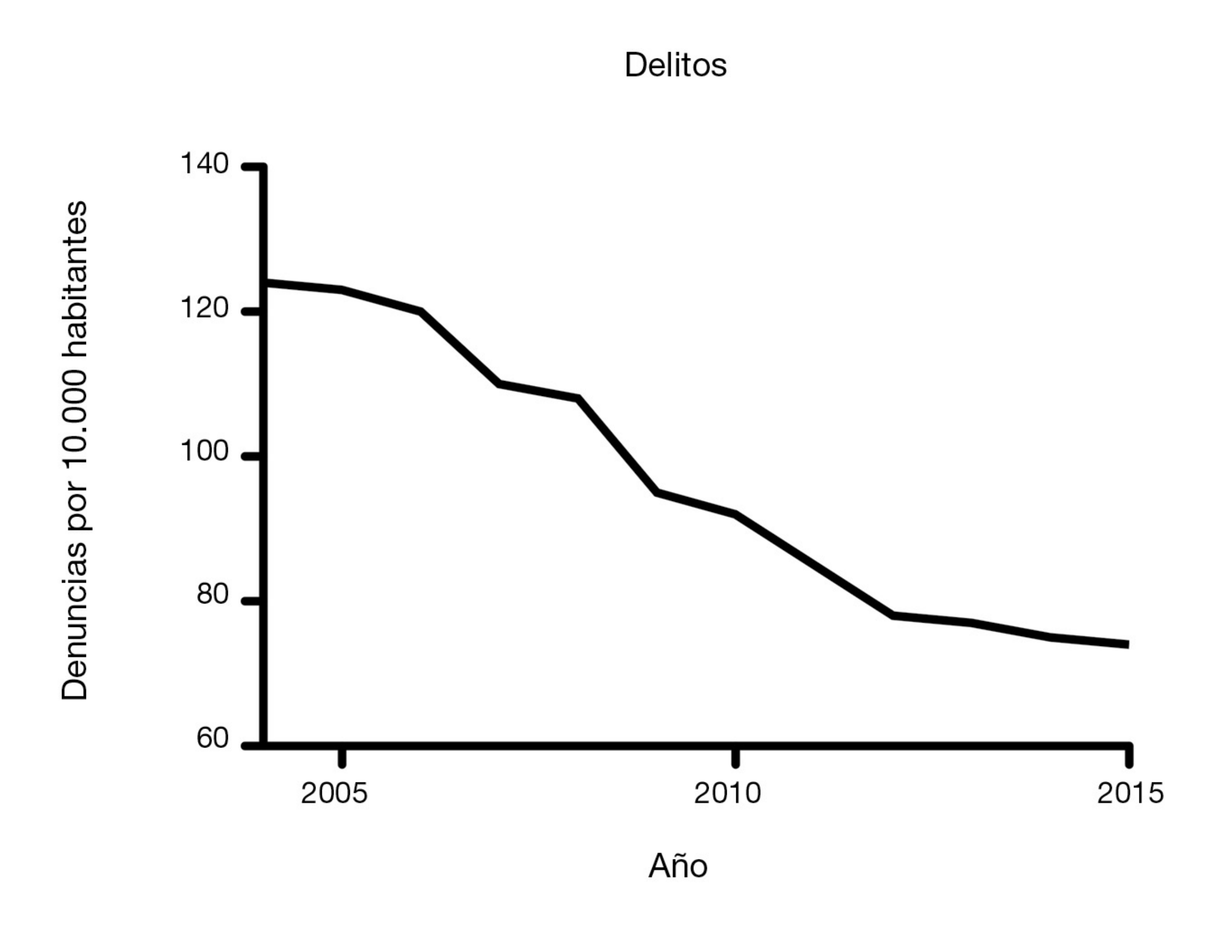
La conducta honesta sería mostrar el primer gráfico de delitos presentándolo con los ejes sin discontinuidades. Pues bien, para poder evaluar críticamente la estadística, debiéramos preguntarnos si hay algún factor que podría estar ocultando la verdad en el procedimiento mediante el que se recopilan o se presentan los datos.

Una de las posibilidades es que los incrementos tengan lugar en un barrio particularmente malo y que, de hecho, los delitos estén *disminuyendo* en los restantes barrios de la ciudad. Quizá la policía y la comunidad han decidido que ese barrio en particular es incontrolable y han dejado de obligar al cumplimiento de la ley allí. La ciudad en su conjunto es segura, incluso puede que sea

más segura que antes, lo que pasa es que un mal barrio es responsable del incremento.

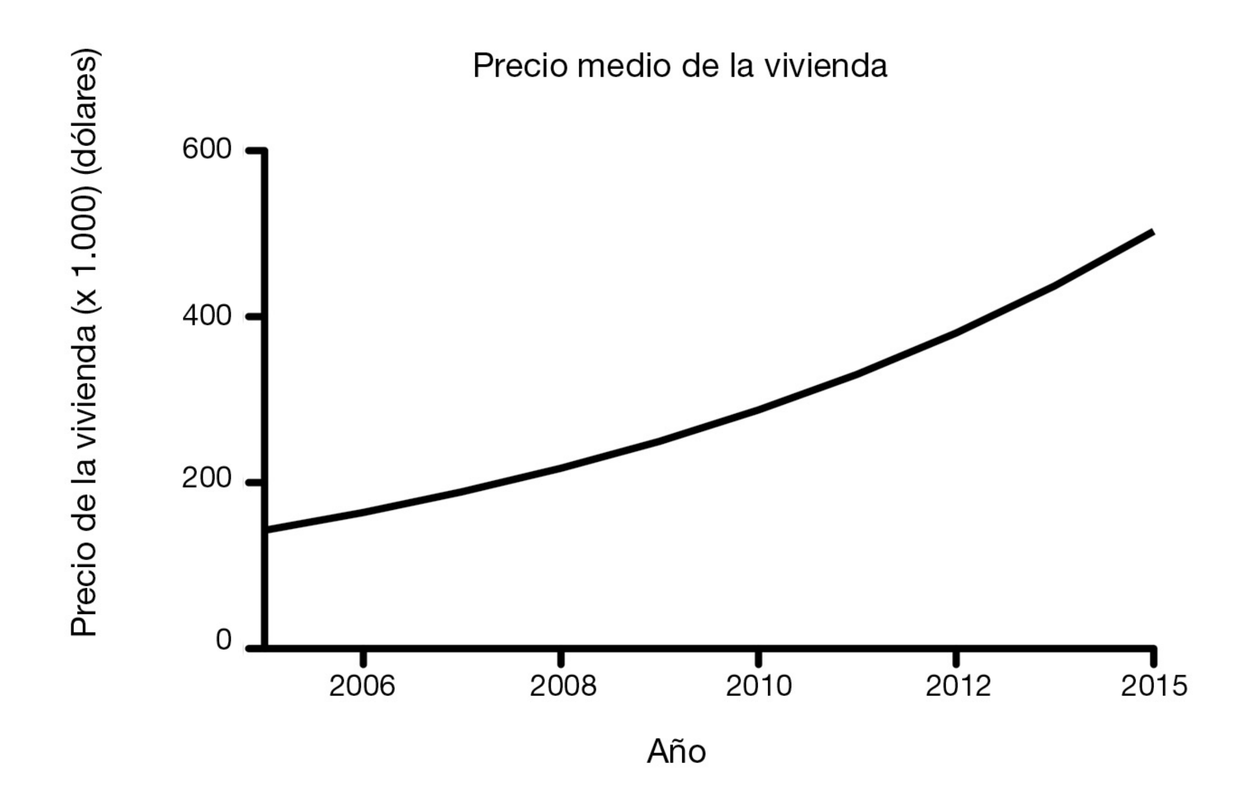
Otra posibilidad es que al agregar todos los diferentes tipos de denuncias en la cesta de los *delitos,* estemos pasando por alto algún aspecto importante. Tal vez los *delitos violentos* han bajado casi hasta cero y, en lugar de éstos y disponiendo de tanto tiempo libre, la policía se dedica a multar a quienes cruzan la calle con el semáforo en rojo.

Quizá la pregunta más obvia que debamos formular seguidamente para entender lo que significa esta estadística es «¿qué ha sucedido con la *población total* de la ciudad durante ese periodo?». Si la población ha aumentado más de un 5% al año, en realidad la tasa de delitos ha disminuido en términos de la proporción por persona. Podemos mostrarlo agregando los delitos cometidos por cada diez mil personas en la ciudad:

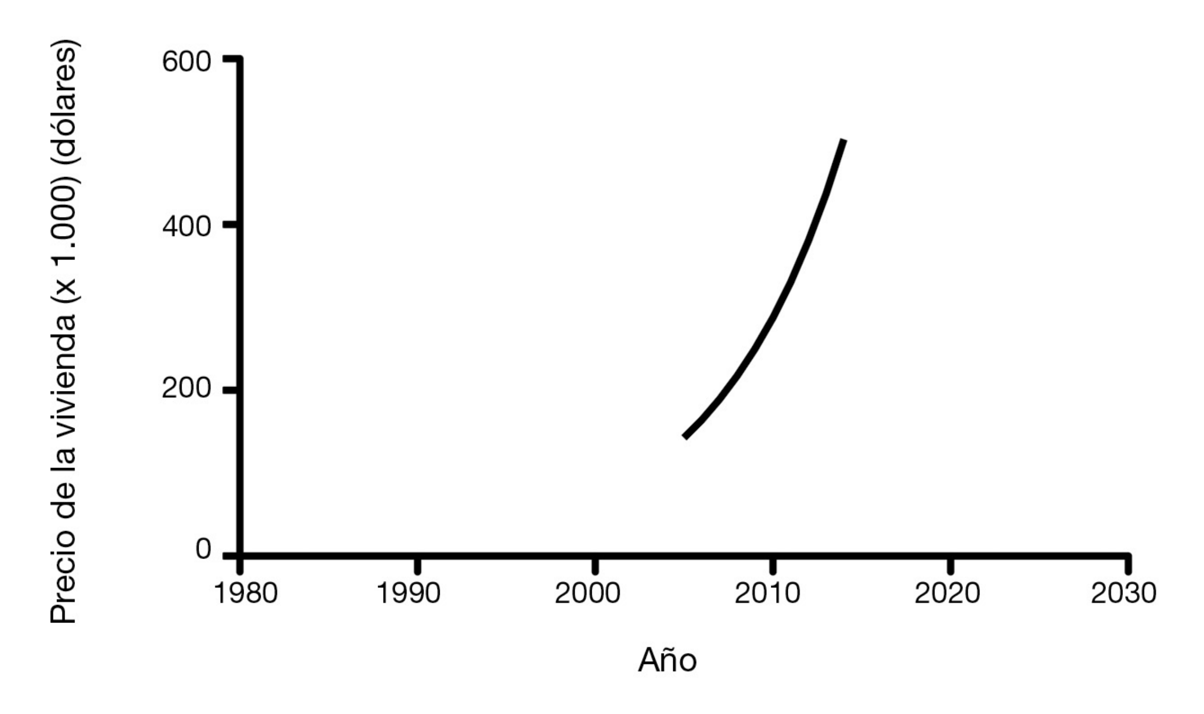


Elección de la escala y ejes adecuados[38]

Habéis sido contratados por una inmobiliaria local para ilustrar los cambios en los precios de la vivienda en vuestra comunidad en la última década. Los precios ha crecido de modo estable un 15% al año.

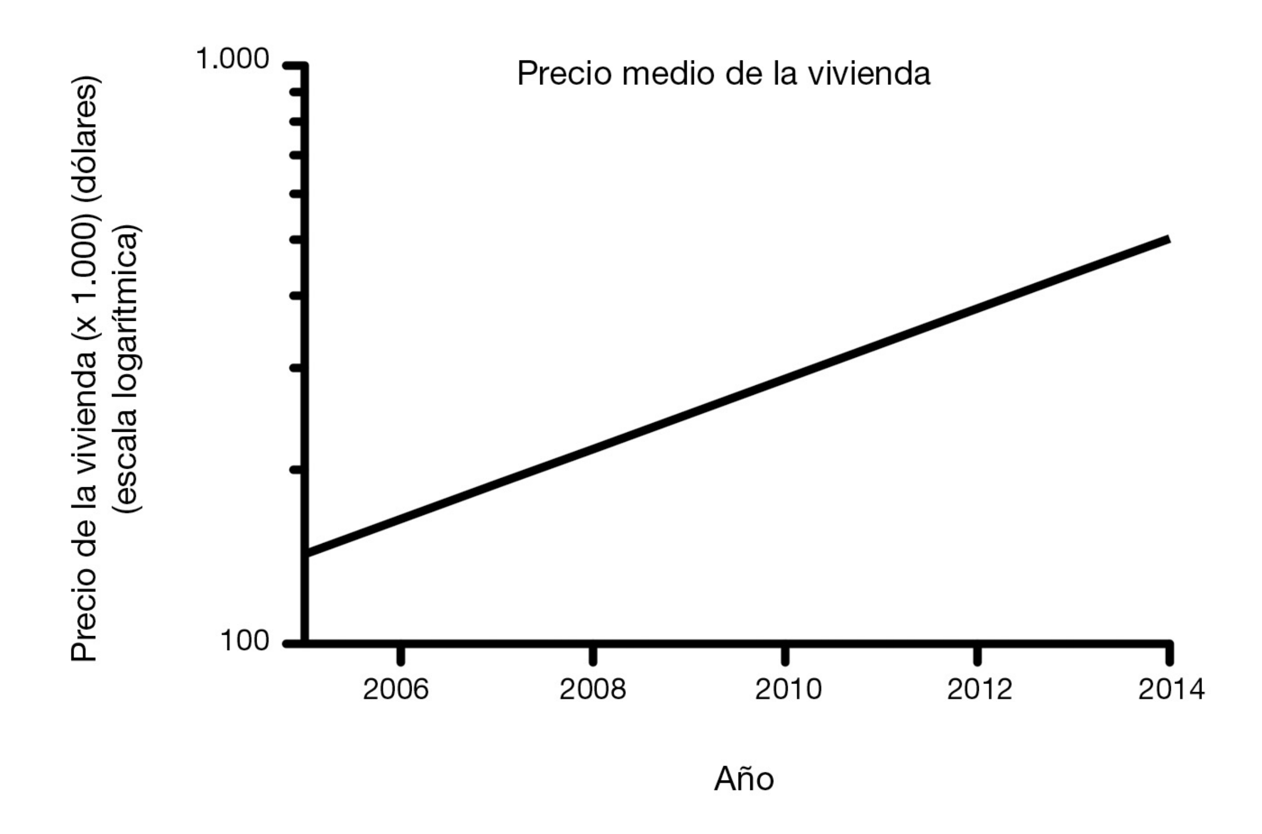


Si lo que queréis es alarmar a la gente, ¿por qué no cambiar el eje *x* para que incluya fechas nuevas sobre las que no tenéis datos? Añadir fechas nuevas artificialmente al eje *x* provoca un incremento de la pendiente de la curva al comprimir la porción visible de la misma, así:



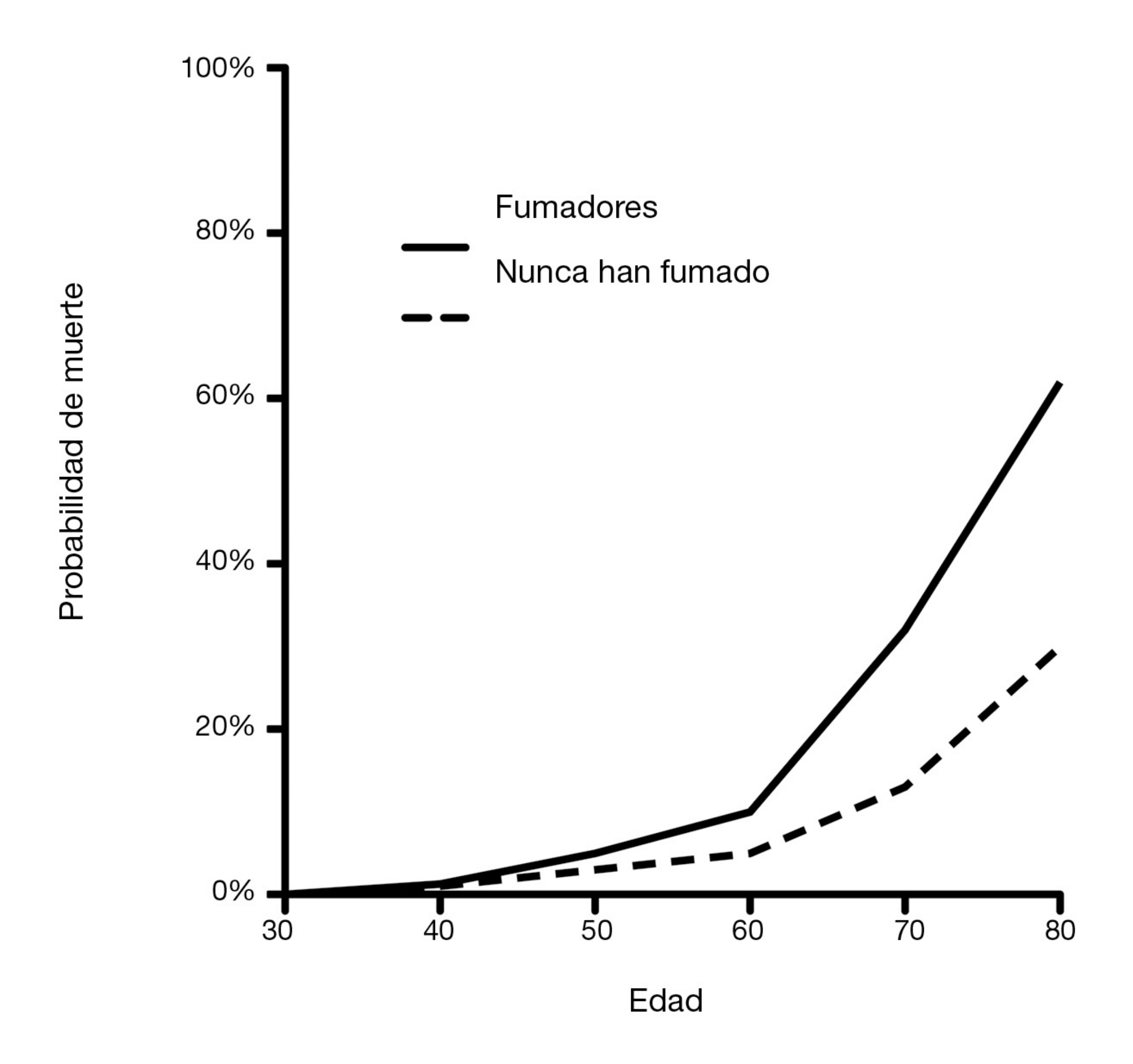
Obsérvese cómo engaña este gráfico a la vista (bueno, al cerebro), arrastrándonos a dos conclusiones falsas; la primera es que en algún momento hacia 1990 los precios de la vivienda debían ser muy bajos, y la segunda es que hacia 2030 los precios de la vivienda serán tan elevados que muy pocas personas podrán comprar una. ¡Mejor será que compréis una ahora mismo!

Ambos gráficos distorsionan la realidad, ya que logran que lo que en realidad es una tasa de incremento estable aparezca visualmente como una tasa de incremento creciente. En el primer gráfico, un 15% de incremento aparece como el doble en el eje *y* en 2014 respecto a 2006. Muchas cosas cambian con una tasa constante[39]: los salarios, los precios, la inflación, la población de una especie, las víctimas de enfermedades. Cuando existe una situación de crecimiento (o de disminución) estable, el modo más preciso de representarla es mediante una escala logarítmica. La escala logarítmica permite que porcentajes de cambio iguales sean representados mediante distancias iguales en el eje *y*. Una tasa de cambio anual constante aparece como una línea recta, así:



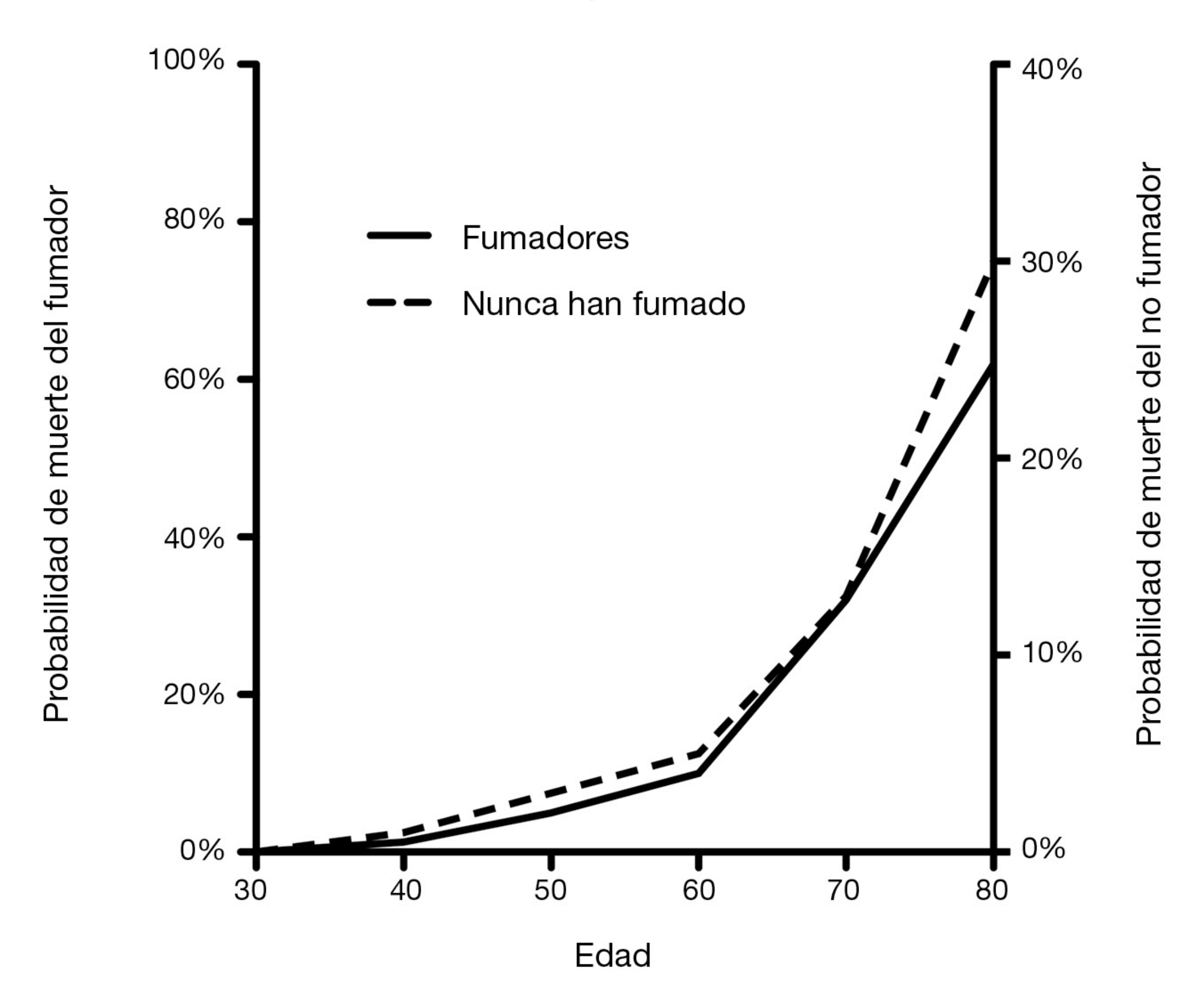
El temible eje y doble

El diseñador de gráficos puede escamotear todo tipo de mentiras sabiendo simplemente que la mayoría de los lectores no van a observar con atención los gráficos. Esto puede llevar a muchas personas a creer todo tipo de patrañas. Consideremos el siguiente gráfico, que muestra la esperanza de vida de fumadores y no fumadores a los veinticinco años de edad[40].

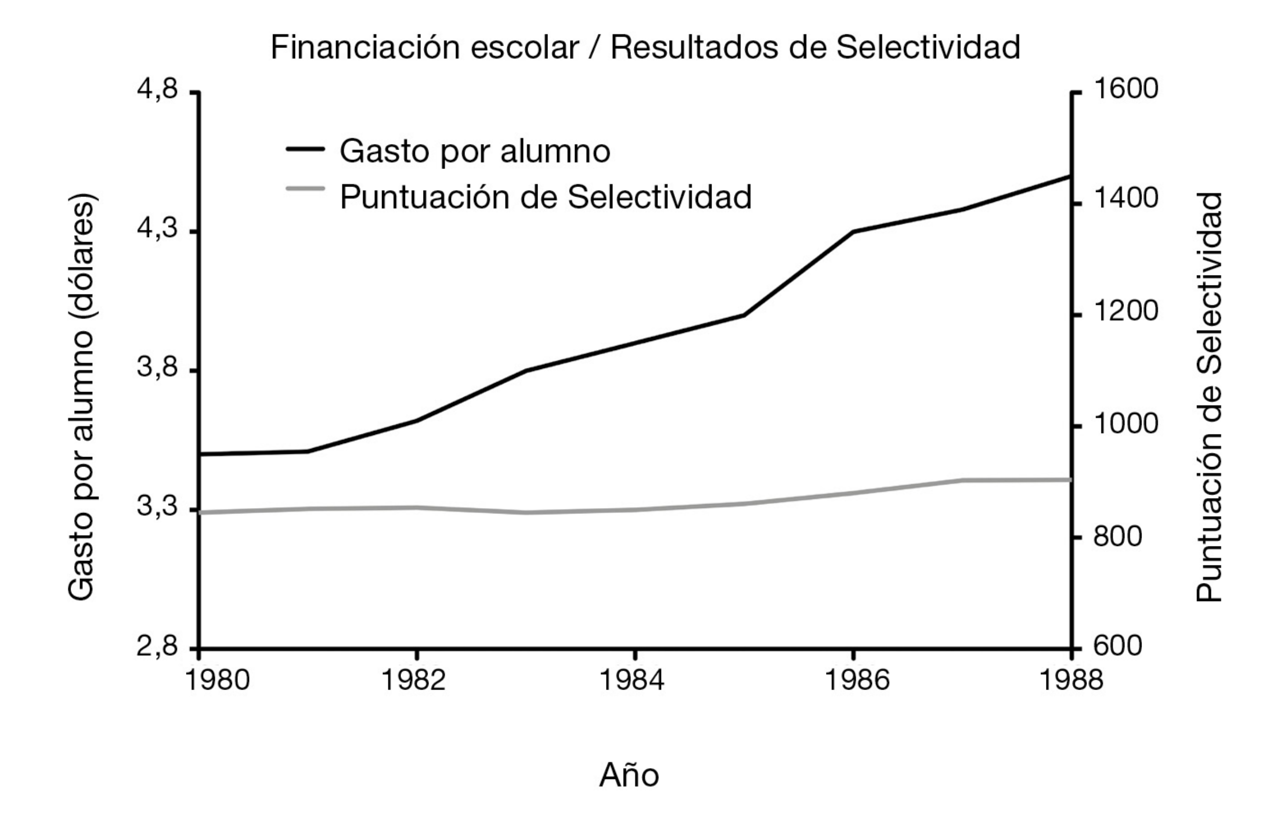


Esto deja claras dos cosas: los peligros de fumar se acumulan con el tiempo y los fumadores tienen más probabilidad de morir más jóvenes. La diferencia no es grande a los cuarenta años, pero hacia los ochenta el riesgo es más del doble, pasando del 30% al 60%. Este es un modo riguroso y claro de presentar los datos. Pero supongamos que sois fumadores de catorce años decididos a convencer a vuestros padres de que deben permitiros fumar. El gráfico tal como está no va a ayudaros mucho. Así que rebuscáis en vuestra bolsa de trucos y utilizáis el eje *y* doble, añadiendo un eje *y* a la derecha del gráfico, con una escala diferente que se aplica únicamente a los no fumadores. Una vez aplicado, vuestro gráfico aparece así:



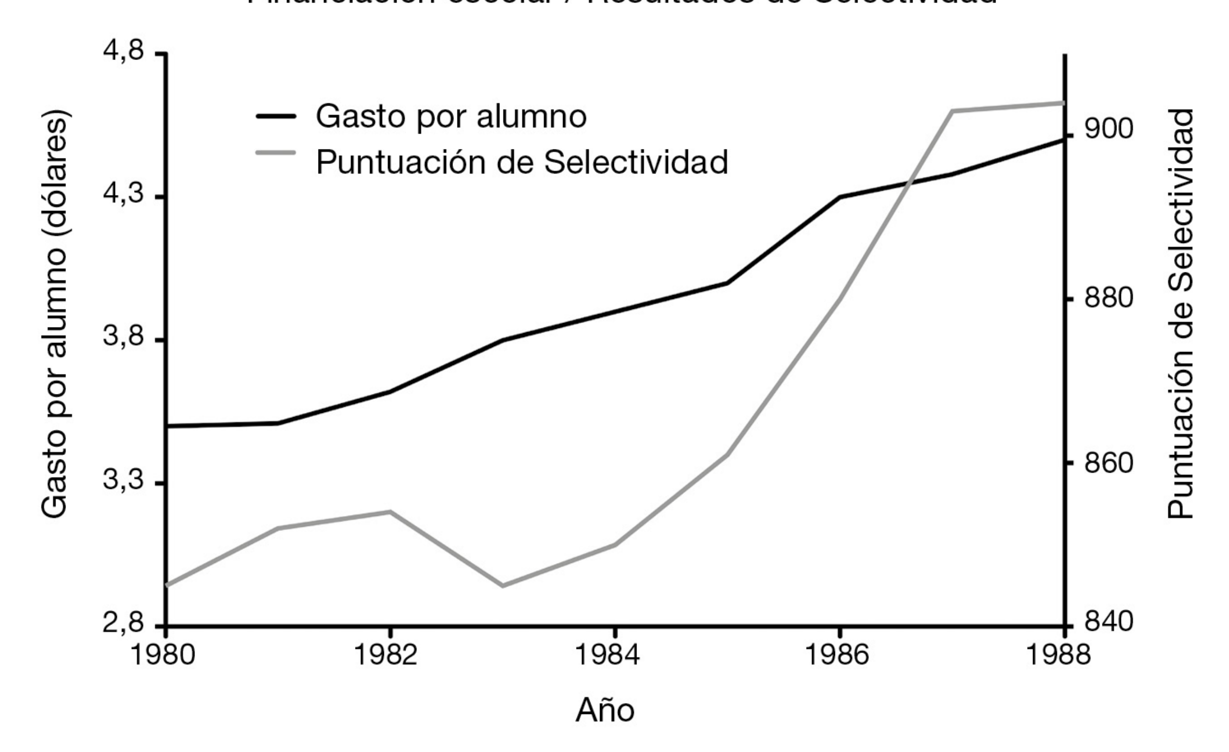


Según parece, tenéis aproximadamente la misma probabilidad de morir por fumar que por no fumar. ¡Fumar no va a haceros daño, son los años quienes se encargarán de eso! El problema de los gráficos con eje *y* doble es que se puede escalar el segundo eje como uno prefiera.



La revista *Forbes,* una fuente de noticias venerable y de confianza, mostró un gráfico muy semejante a este[41] para ilustrar la relación entre el gasto por alumno en las escuelas públicas y las notas de Selectividad (SAT)[42], una prueba estandarizada que se utiliza mucho en Estados Unidos para el acceso a la universidad.

Según el gráfico, da la impresión de que incrementar el dinero que se gasta por alumno (línea negra) no produce un aumento en sus puntuaciones de Selectividad (línea gris). Algunos políticos contrarios al gasto público utilizarían esta historia para afirmar que se despilfarra el dinero de los contribuyentes. Pero ahora sabemos que la elección de la escala del segundo eje *y* (a la derecha) es arbitraria. Si fuéramos administradores escolares podríamos utilizar exactamente los mismos datos y *¡voilà!* Un incremento en los gastos da lugar a una educación mejor, ¡tal como se muestra por el aumento de las puntuaciones de Selectividad!

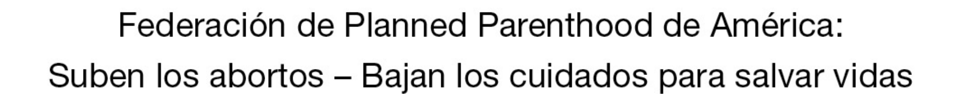


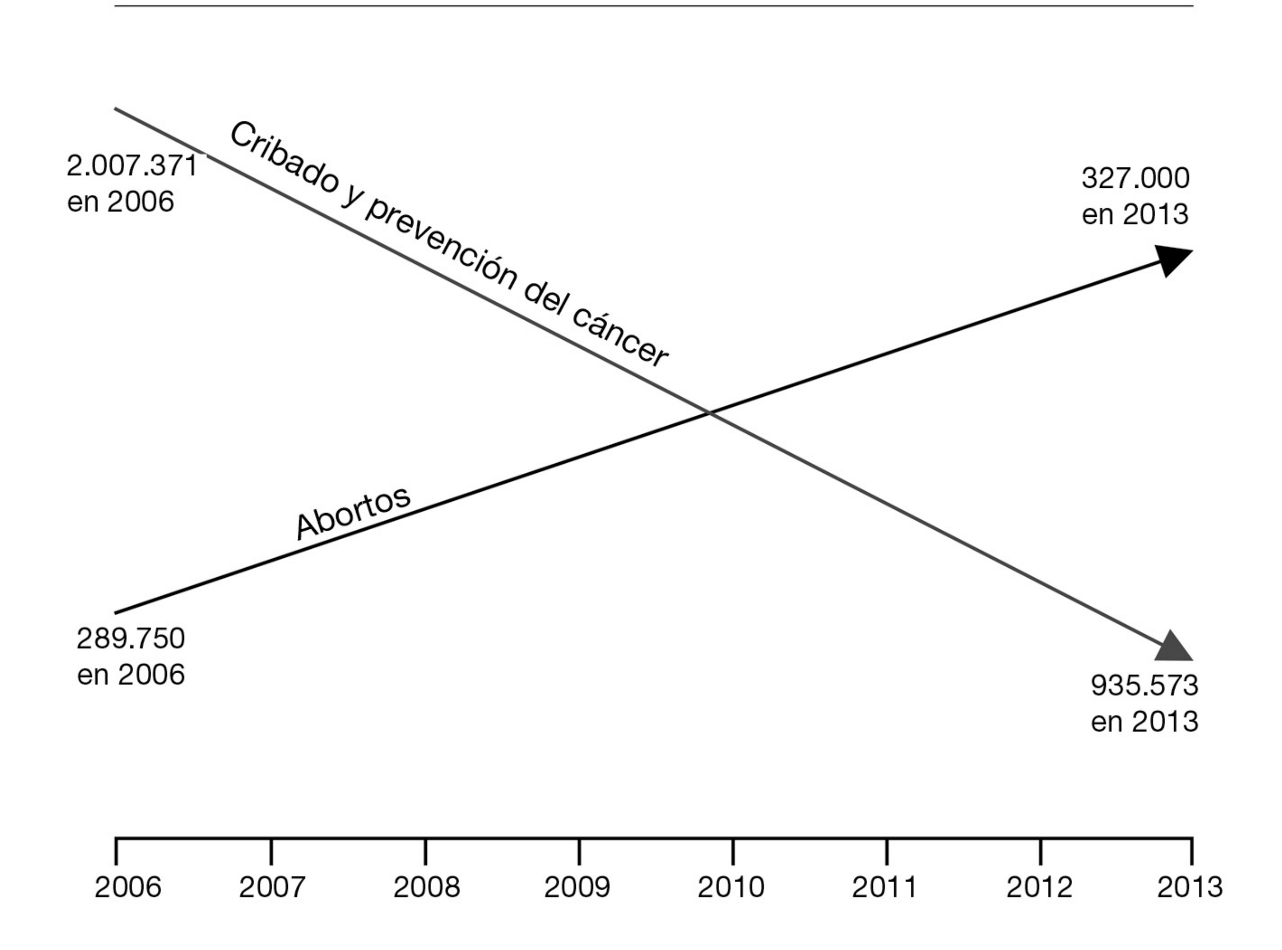
Está claro que este gráfico nos cuenta una historia bastante diferente. ¿Dónde está la verdad? Sería necesario disponer de una medida sobre cómo cambia una variable en función de la otra, un estadístico denominado correlación. Las correlaciones varían de –1 a 1. Una correlación de 0 indica que una variable no está relacionada con la otra. Una correlación de –1 significa que cuando una variable aumenta, la otra disminuye de modo sincronizado. Una correlación de 1 significa que cuando una variable sube, la otra también lo hace de modo sincronizado. El primer gráfico parece ilustrar una correlación de 0, el segundo gráfico parece representar una próxima a 1. La correlación real para ese conjunto de datos es de 0,91, una correlación muy intensa. Gastar más por cada alumno, al menos según ese conjunto de datos, está asociado con mejores puntuaciones de Selectividad.

La correlación también proporciona una buena estimación[43] de hasta qué punto un resultado puede explicarse mediante las variables sometidas a consideración. La correlación de 0,91 nos indica que podemos explicar el 91% de las puntuaciones de los estudiantes en la Selectividad observando el gasto por alumno de

la escuela. Es decir, nos informa sobre la medida en que los gastos pueden explicar la diversidad de las puntuaciones de Selectividad.

En el otoño de 2015 surgió una controversia sobre el gráfico de *y* doble durante una reunión de un comité del Congreso de los Estados Unidos. El congresista Jason Chaffetz presentó un gráfico que agregaba dos servicios que presta la organización Planned Parenthood[44]: el aborto y el cribado y prevención del cáncer[45]:

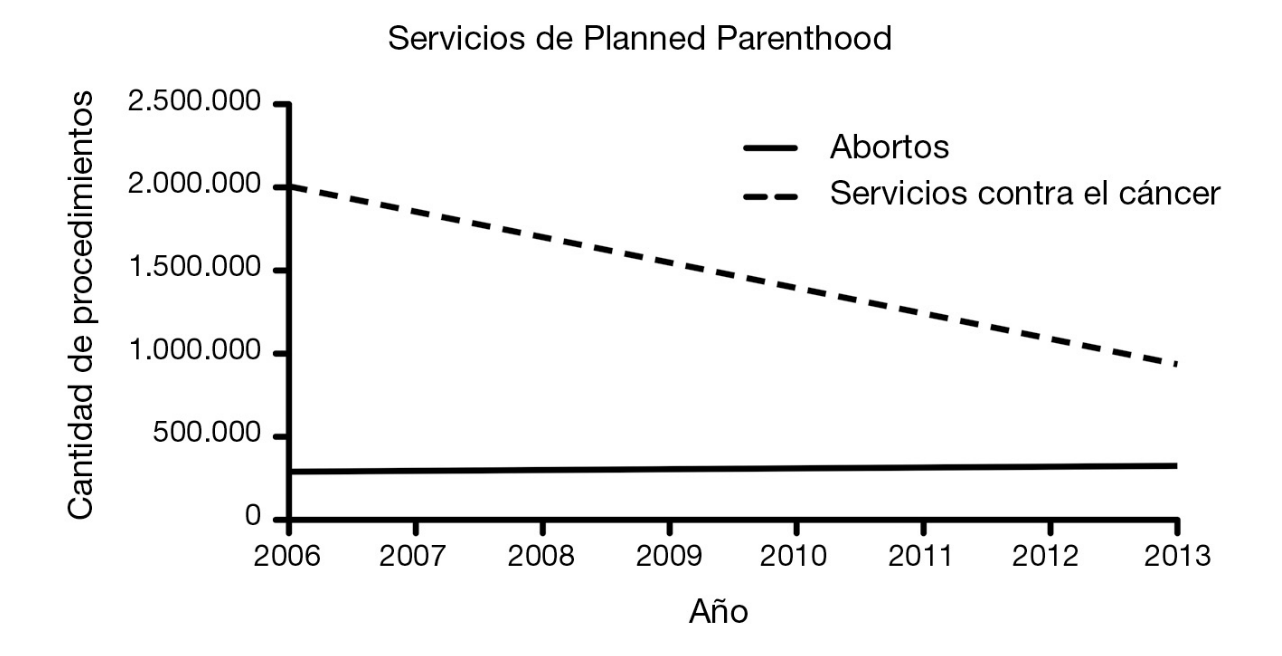




El congresista estaba intentando transmitir un mensaje político: que durante un periodo de siete años, Planned Parenthood había incrementado la cantidad de abortos que había realizado (algo a lo que Chaffetz se opone) y disminuido la cifra de cribados y procedimientos de prevención del cáncer. Planned Parenthood no

lo niega, pero ese gráfico distorsionado sugiere que la cantidad de abortos superó a los procedimientos contra el cáncer. Tal vez quien elaboró el gráfico se sentía un poco culpable, ya que incluyó los datos reales al lado de sus marcas. Vamos a aceptar sus retazos de información y a observarlos atentamente. La cifra de abortos en 2013, el año más próximo para el que disponemos de datos, es de 327.000. La cifra de procedimientos contra el cáncer fue de casi el triple, 935.573. (Por cierto, resulta sospechoso que los datos sobre abortos sean una cifra redonda, en tanto que sobre el cáncer sean exactos). Se trata de un ejemplo particularmente siniestro: ¡un gráfico de *y* doble implícito sin ejes a ninguno de los lados!

Si se hubiera trazado correctamente el gráfico, este sería el siguiente:



Ahora podemos observar que los abortos se incrementaron modestamente comparados con la reducción de los servicios contra el cáncer.

Hay otro aspecto sospechoso en el gráfico original: esas líneas regulares raramente se encuentran en los datos. Parece más probable que quien elaboró el gráfico se limitó a utilizar los datos de dos años en concreto, 2006 y 2013, y los comparó, dibujando una línea recta regular entre ellos. Tal vez esos años fueron

elegidos para resaltar intencionadamente las diferencias. Quizá hubo grandes fluctuaciones en los años intermedios, entre 2007 y 2012, no lo sabemos. Las líneas regulares producen la impresión de una función perfectamente lineal (una línea recta), lo que resulta muy improbable.

Gráficos como el citado no siempre describen lo que la gente cree. ¿Hay algo que podría explicar estos datos, aparte de la historia de que Planned Parenthood está embarcado en la tarea de realizar todos los abortos que pueda (y dejar que la gente muera de cáncer al mismo tiempo)? Observemos el segundo gráfico. En 2006, Planned Parenthood realizó 2.007.371 procedimientos contra el cáncer y 289.750 abortos: casi siete veces más procedimientos contra el cáncer que abortos. En 2013 esa diferencia había disminuido, pero la cantidad de servicios contra el cáncer continuaba siendo casi el triple que la cifra de abortos.

Cecile Richards, presidente de Planned Parenthood ofreció una explicación sobre esa diferencia menguante. El cambio en algunas de las normas sanitarias en los servicios de lucha contra el cáncer, como el papanicolau, redujo la cifra de personas a quienes se recomendaba esa supervisión. Otros cambios, como las actitudes sociales hacia el aborto, los cambios de edad en la población y el aumento del acceso a otras alternativas de cuidados sanitarios, influyen sobre estas cifras, de modo que los datos que se presentaron no demuestran que Planned Parenthood tenga un programa proaborto; podría ser, pero estos datos no lo prueban.

---

33 L. M. Tully, S. H. Lincoln, T. Wright y C. I. Hooker (2013), «Neural mechanisms supporting the cognitive control of emotional information in schizophrenia», póster presentado en la 25th Annual Meeting of the Society for Research in Psychopathology, https://www.researchgate.net/publication/266159520_Neural_mechanisms_supporting_the_cognitive_control_of_emotional_information_in_schizophrenia. Encontré este ejemplo en www.betterposters.blogspot.com.

34 http://pelgranepress.com/index.php/tag/biz/.

35 *N. del T.:* Kickstarter es una gran plataforma de recaudación de financiación para proyectos creativos.

36 Lo he redibujado a fin de aclararlo. El original puede consultarse en http://cloudfront.mediamatters.org/static/images/item/fbn-cavuto-20120731-bushexpire.jpg.

37 Spirer, Spirer y Jaffe, *op. cit.,* pp. 82-84.

38 Ejemplo de Spirer, Spirer y Jaffe, *op. cit.,* p. 78.

39 Spirer, Spirer y Jaffe, *op. cit.,* p. 78.

40 Estos datos se han tomado de P. Jha *et al.* (2013), «21st-century hazards of smoking and benefits of cessation in the United States», *New England Journal of Medicine,* 368(4), 341-350, figura A2 para las mujeres. Las probabilidades de supervivencia se han adaptado de las tasas de mortalidad por todas las causas a estas edades para el año 2004, ajustando las diferencias en edad, nivel educativo, consumo de alcohol y adiposidad (índice de masa corporal) del National Health Interview Survey de los Estados Unidos. Agradezco a Prabhat Jha su correspondencia para la interpretación de estos datos. Esta forma de presentación se basa en la de H. Wainer (1997), *Visual Revelations: Graphical Tales of Fate and Deception from Napoleon Bonaparte to Ross Perot,* Nueva York, Copernicus/Springer-Verlag.

41 Este ejemplo se ha tomado de H. Wainer (1997), *Visual Revelations: Graphical Tales of Fate and Deception from Napoleon Bonaparte to Ross Perot,* Nueva York, Copernicus/Springer-Verlag, p. 93. El original apareció en *Forbes* (14 de mayo de 1990). Sin duda concurren otras variables. Los gastos de los que se informa, ¿están expresados en dólares reales o ajustados a la inflación? ¿Se escogió el periodo 1980-1988 para poder resaltar una conclusión y en otro periodo se hubiera alcanzado otra diferente?

42 *N. del T.:* Traducimos como «Selectividad» la prueba SAT (Scholastic Aptitude Test), una prueba que frecuentemente se emplea en Estados Unidos para regular el acceso a muchas universidades.

43 Existe una cierta controversia sobre si debe utilizarse *r* o *r-cuadrado*. Para la defensa de *r* véase: R. D'Andrade y J. Dart (1990), «The interpretation of r versus r2 or why percent of variance accounted for is a por measure of size effect», *Journal of Quantitative Anthropology,* 2, 47-59. D. J. Ozer (1985), «Correlation and the coefficient of determination», *Psychological Bulletin,* 97(2), 307-315.

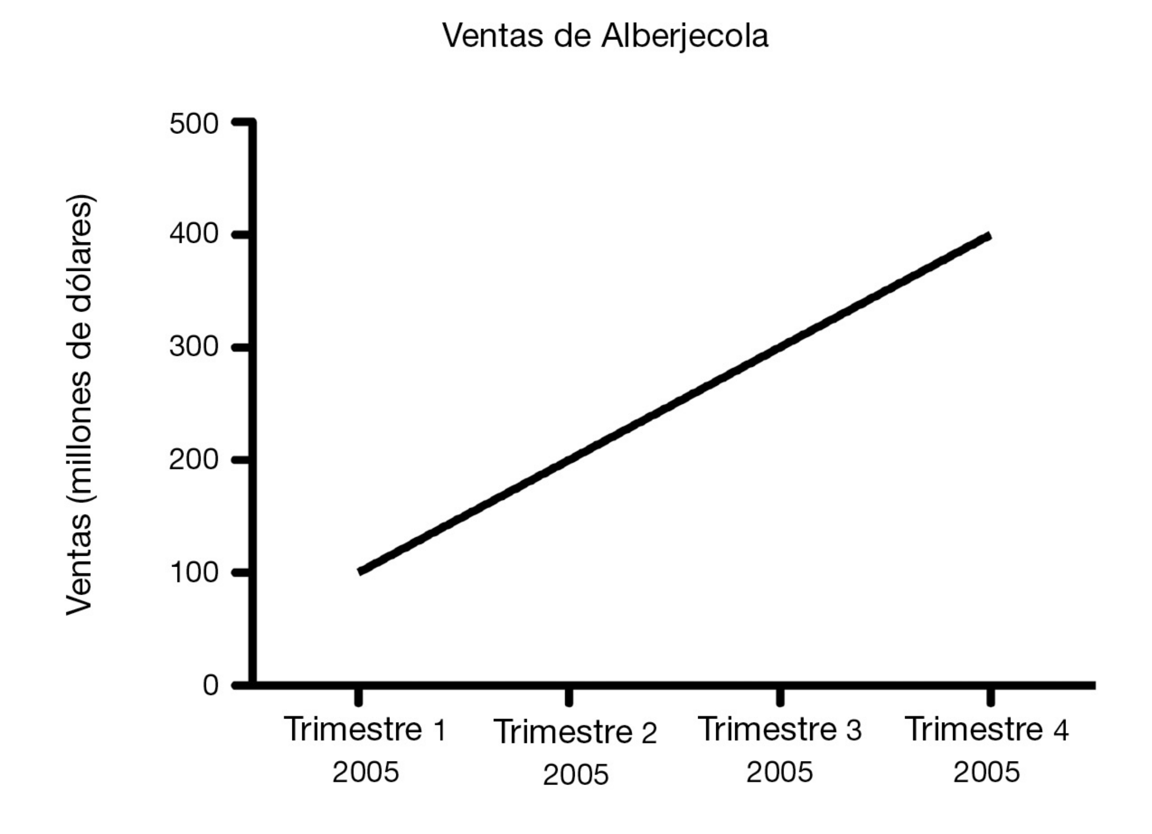
44 *N. del T.:* Planned Parenthood es una conocida organización dedicada a promover la salud sexual y reproductiva.

45 Z. Roth (29 de septiembre de 2015), «Congressman uses misleading graph tos mear Planned Parenthood», msnbc.com. Politifact exploró más a fondo este asunto, examinando los datos entre los dos momentos y proporcionó

información contextual adicional acorde con la habitual crítica centrada en el gráfico. Véase https://perma.cc/P8NY-YP49.

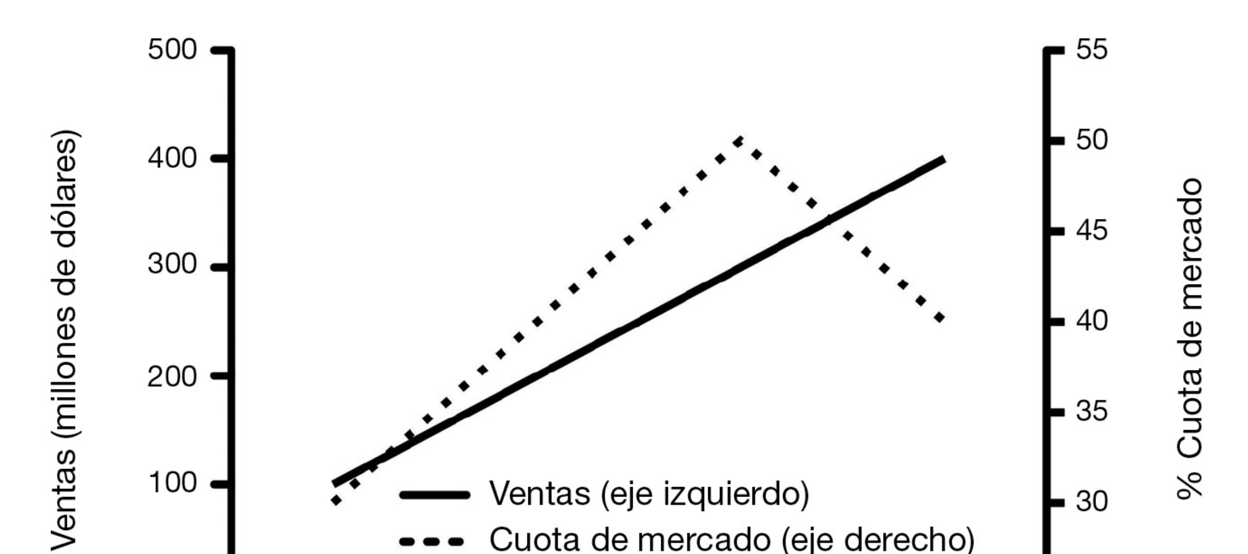
PICARDÍAS AL DESCRIBIR LAS CIFRAS

Estamos intentando decidir si hay que comprar un lote de un refresco nuevo y nos topamos con este gráfico sobre las cifras de ventas del fabricante en su informe anual:



Los datos parecen prometedores, Alberjecola incrementa sus ventas de modo estable. Por ahora todo bien, pero apliquemos un poco de sentido común. El mercado de los refrescos es muy competitivo. Las ventas de Alberjecola están aumentando, pero podría no ser tan rápidamente como las de sus competidores. Como inversores potenciales lo que realmente necesitamos conocer es la comparación entre las ventas de Alberjecola y las de los demás fabricantes, o conocer su cuota de mercado, las ventas de Alberjecola podrían estar aumentando ligeramente, en tanto que el mercado crece mucho y los competidores obtienen

más beneficios que los fabricantes de Alberjecola. Como demuestra este ejemplo de un útil gráfico de doble eje *y,* esto podría no ser un buen indicio para el futuro:



Pese a que los diseñadores de gráficos carentes de escrúpulos puedan hacer artimañas con la escala del eje derecho a fin de conseguir que el gráfico muestre cualquier cosa que deseen, este tipo de gráfico de doble eje *y* no es ultrajante, ya que los dos ejes *y* representan cosas diferentes, cantidades que *no pueden* compartir un eje. Es un caso diferente del gráfico de Planned Parenthood de la página 59, que daba cuenta de la misma cantidad en dos ejes distintos: la cantidad de procedimientos que se habían llevado a cabo. Ese gráfico implicaba una distorsión, ya que aunque los dos ejes medían lo mismo, su escala era diferente con la intención de manipular nuestra percepción.

También resultaría útil considerar los beneficios de Alberje: quizá mediante el empleo de reducciones de coste en la fabricación y la distribución podrían estar obteniendo mayores beneficios con un volumen de ventas menor. El mero hecho de que alguien cite una estadística o nos muestre un gráfico no significa que sea pertinente para el asunto que se pretende

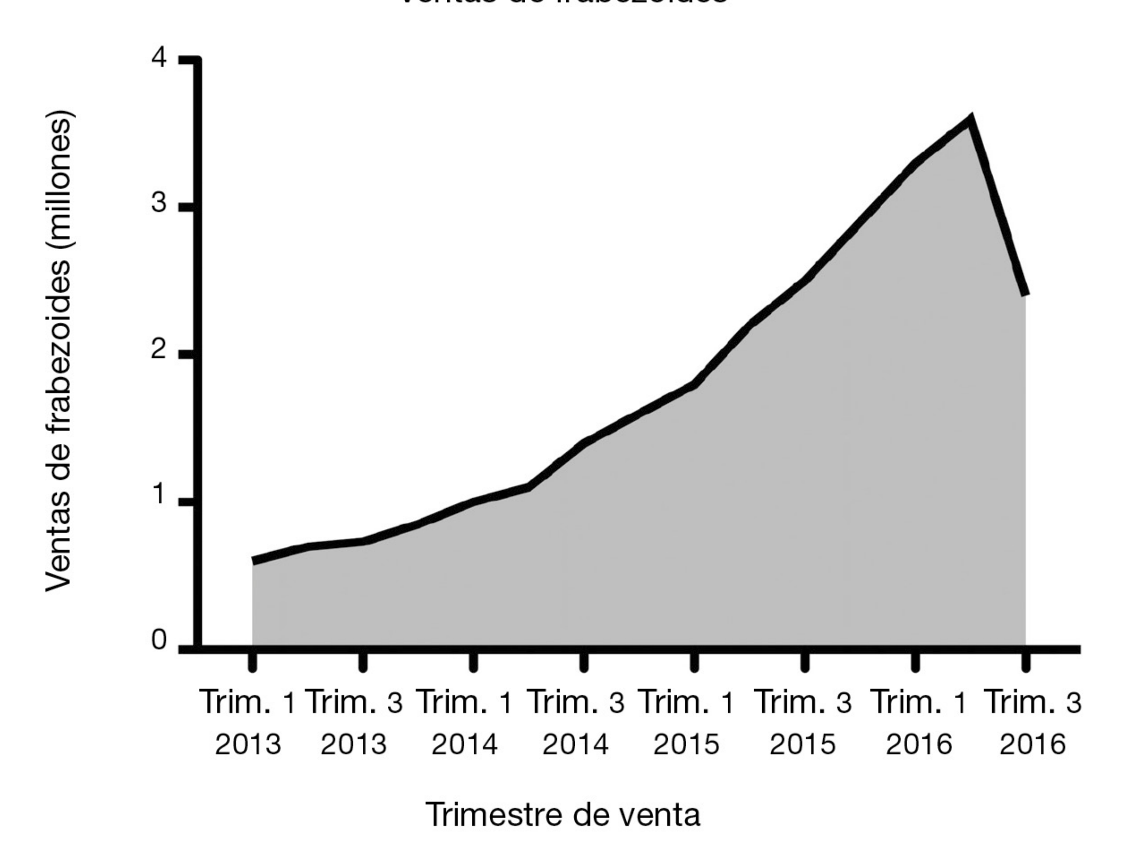
probar. Nuestro deber es asegurarnos de obtener la información relevante e ignorar la información que no lo es.

Supongamos que estáis trabajando en el departamento de relaciones públicas de una empresa que fabrica algún aparato, frabezoides, por ejemplo. Durante los últimos años el ansia del público por los frabezoides ha sido grande y las ventas se han incrementado. La empresa ha crecido y ha construido fábricas nuevas, ha contratado más empleados y ha elevado los salarios. Vuestro jefe entra en vuestro despacho con una expresión sombría y os explica que han llegado los datos de ventas más recientes y las ventas de los frabezoides han caído un 12% en comparación con el trimestre anterior. El presidente de la empresa está a punto de ofrecer una gran conferencia de prensa para hablar del futuro de la empresa. Como es costumbre, va a mostrar un gráfico enorme en la pantalla a su espalda, mostrando cómo marchan los frabezoides. Si llegara a conocerse la caída en la venta de frabezoides, el público podría pensar que ya no son objetos deseables y eso podría provocar una caída de las ventas aún mayor.

¿Qué vais a hacer? Si hacéis un gráfico que represente honestamente los datos de ventas de los cuatro últimos años, el gráfico sería como el primero de la página siguiente.

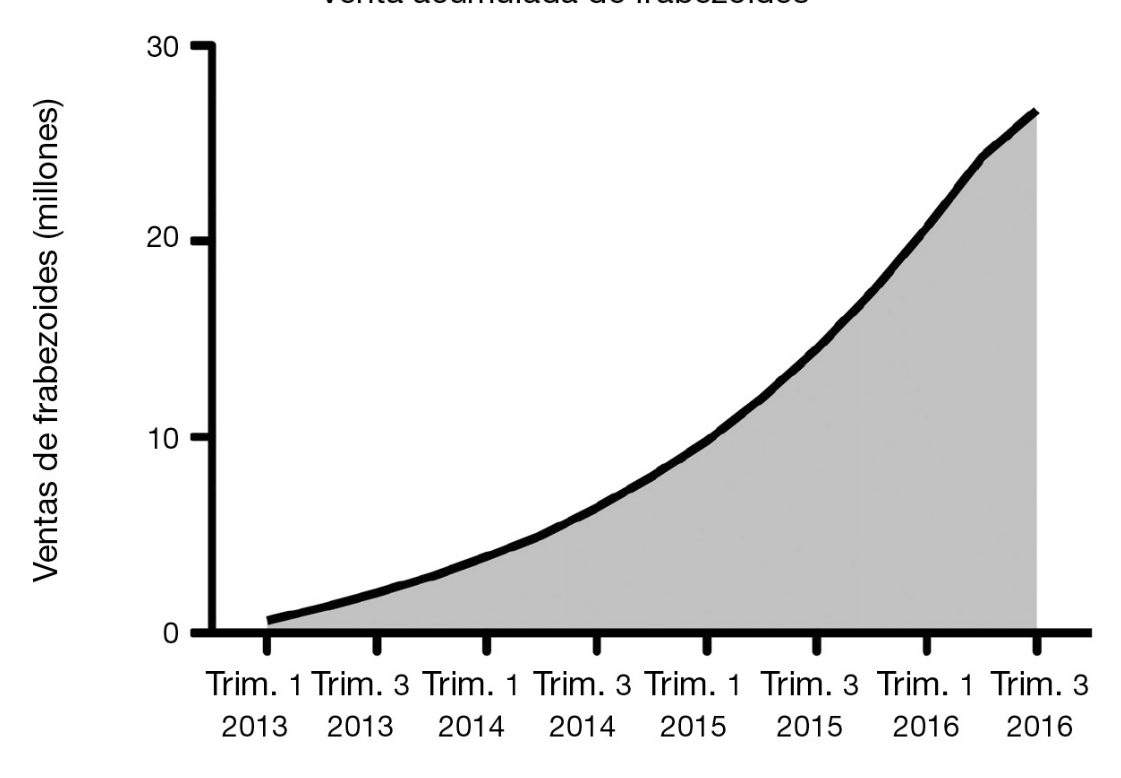
El problema es esa tendencia a la baja en la curva. Si hubiera algún medio de hacer que la curva subiera...

Pues bien, ¡existe! El gráfico de ventas acumulativo. En lugar de representar las ventas por trimestres, representad las ventas acumuladas por trimestre, o sea, el total de ventas hasta la fecha.



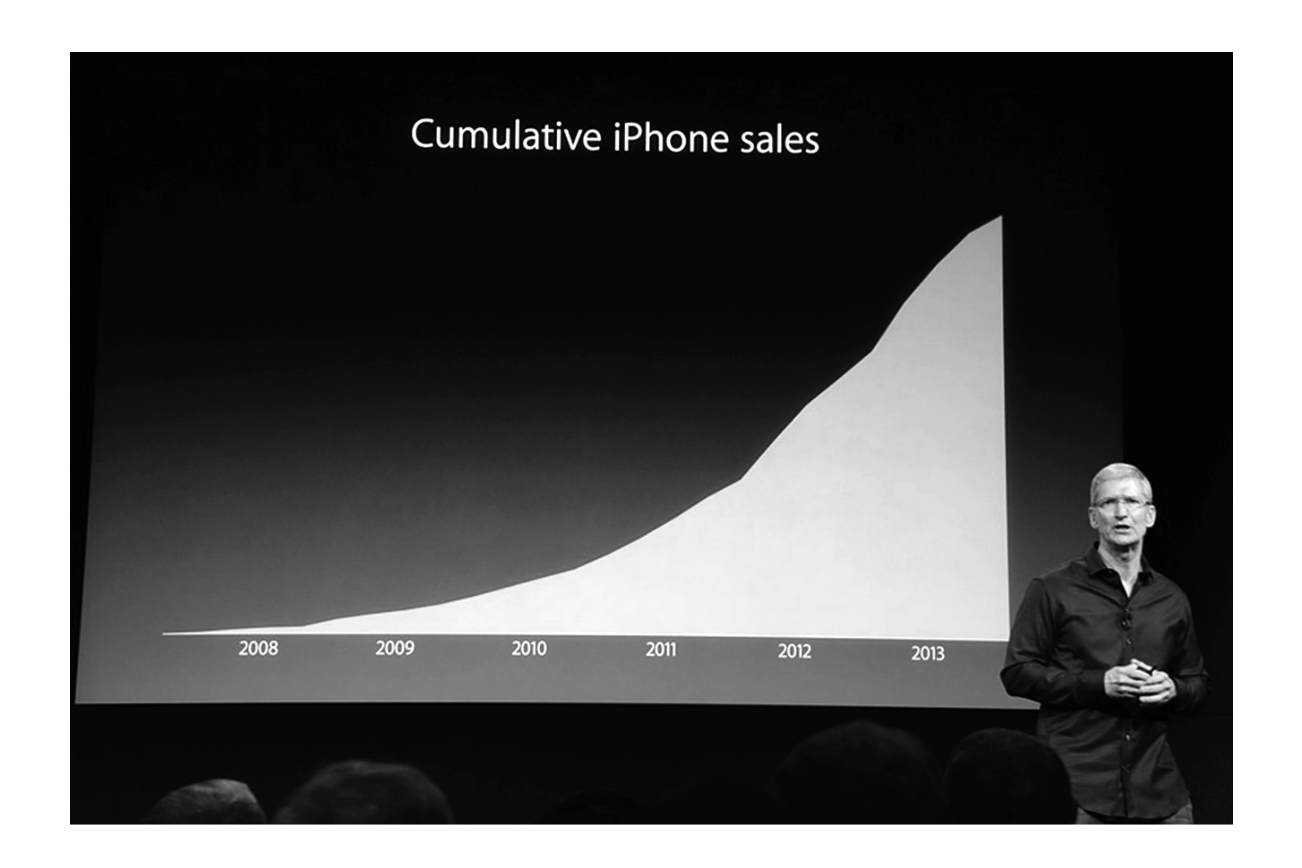
Mientras hayáis logrado vender un solo frabezoide, vuestro gráfico acumulativo mostrará un incremento, como este:





Si observáis cuidadosamente, todavía es posible ver un vestigio de las malas ventas del último trimestre: pese a que la línea continúa ascendiendo, lo hace de un modo menos pronunciado. Esa es la pista que indica que las ventas han caído. Pero nuestros cerebros no son muy buenos en la detección de tasas de cambio como esa (lo que se denomina la primera derivada en cálculo, un nombre de fantasía para la inclinación de una línea). De modo que si nos limitamos a una observación superficial, parece que la empresa continúa yendo muy bien y habréis logrado que muchos consumidores continúen creyendo que los frabezoides son justo lo que está de moda.

Esto es exactamente lo que Tim Cook, director ejecutivo de Apple, hizo al presentar las ventas del iPhone[46].



Agregar asuntos que no tienen relación

En este mundo suceden tantas cosas que necesariamente ocurren coincidencias. La cantidad de camiones de color verde en una carretera puede aumentar al mismo tiempo que nuestro sueldo; cuando éramos niños puede que aumentara la cantidad de programas al tiempo que se incrementaba nuestra estatura. Pero eso no significa que lo uno sea la causa de lo otro. Cuando dos sucesos están relacionados, sea o no de modo causal, los estadísticos lo denominan correlación.

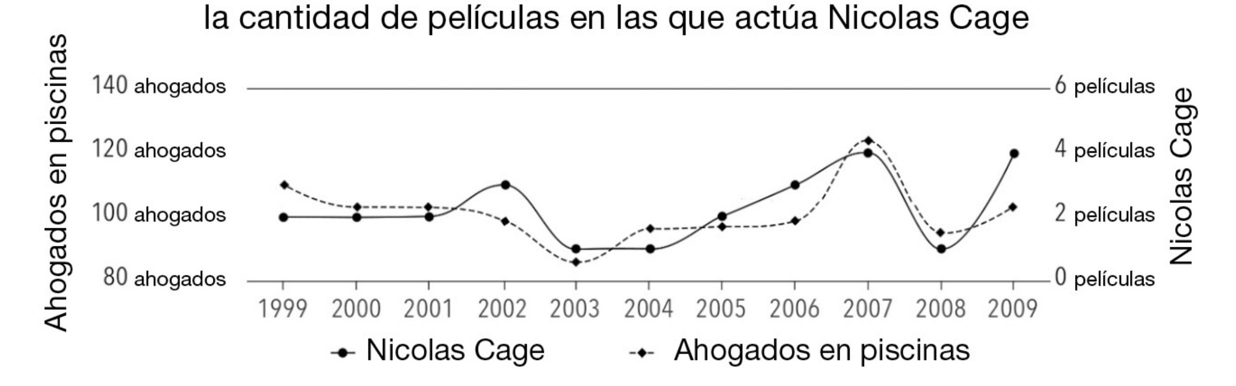
Hay una famosa sentencia que dice que «la correlación no supone causación». En la lógica formal hay dos formulaciones de esta regla:

1) *Post hoc, ergo propter hoc* (tras esto, por tanto a causa de esto). Se trata de una falacia lógica que surge de creer que simplemente porque un suceso (Y) tiene lugar después de otro (X), X ha *causado* Y. Es habitual que las personas se cepillen los dientes por la mañana antes de irse a trabajar, pero cepillarse los dientes no es la *causa* de que vayan a

trabajar. En este caso incluso es posible que se trate de lo inverso.

2) *Cum hoc, ergo propter hoc* (al mismo tiempo, por tanto a causa de esto). Esta es la falacia lógica que surge de creer que simplemente porque dos sucesos ocurren al mismo tiempo, uno ha de ser la causa del otro. Para dejar bien claro este asunto, el alumno de la Facultad de Derecho de Harvard, Tyler Vigen, ha escrito un libro y desarrollado una web que ilustra coocurrencias espurias[47], correlaciones, como la siguiente:

La cantidad de personas que se han ahogado al caerse a una piscina correlaciona con la cantidad de películas en las que actúa Nicolas Cage



Hay cuatro modos de interpretar esto: (1) los ahogados causan la aparición de películas nuevas de Nicolas Cage; (2) las películas nuevas de Nicolas Cage causan los ahogados; (3) un factor (aún no identificado) causa ambos sucesos; o (4) simplemente no hay relación y la correlación es una coincidencia. Si no distinguimos correlación de causación, podríamos afirmar que el gráfico de Vigen «prueba» que Nic Cage contribuye a la prevención de los ahogamientos en piscinas, es más, lo mejor que podríamos hacer es animarle a hacer menos películas, de modo que pueda ejercer sus habilidades de salvavidas, como parece que hizo con gran provecho en 2003 y 2008.

En algunos casos no hay una conexión real entre los sucesos que correlacionan, su correlación es una mera coincidencia. En otros casos es posible descubrir un vínculo causal entre los

sucesos correlacionados o, al menos, especular con alguna idea razonable que nos lleve a buscar más información.

Podemos desechar la primera explicación, ya que producir y distribuir una película tarda bastante tiempo, de modo que un incremento en los ahogamientos no puede causar un incremento en las películas de Nic Cage ese mismo año. ¿Qué hay de la segunda? Tal vez la gente se queda tan absorta con los dramas de las películas de Cage que se distraen y como consecuencia se ahogan. Podría ocurrir que el mismo ensimismamiento cinematográfico también incremente las tasas de accidentes de automóvil y de los heridos al manipular maquinaria pesada. No lo sabremos hasta que no analicemos más datos, ya que no se nos aportan.

¿Y si un tercer factor causara ambos sucesos? Podríamos sospechar que hay tendencias económicas que provocan ambos hechos: una economía que va mejor promueve más actividades de ocio, se hacen más películas, la gente se va más de vacaciones y van más a nadar. Si esto fuera verdad, entonces ninguno de los hechos ilustrados en el gráfico, las películas de Nic Cage y los ahogamientos, se causaron mutuamente. Por el contrario, fue un tercer factor, la economía, lo que causó ambos sucesos. Los estadísticos llaman a esto el *tercer factor x* de explicación de las correlaciones y hay muchos casos del mismo.

Lo más probable es que ambos sucesos no tengan ninguna relación. Si buscamos el tiempo suficiente y con la atención adecuada, seguro que encontraremos dos sucesos no relacionados que varían juntos.

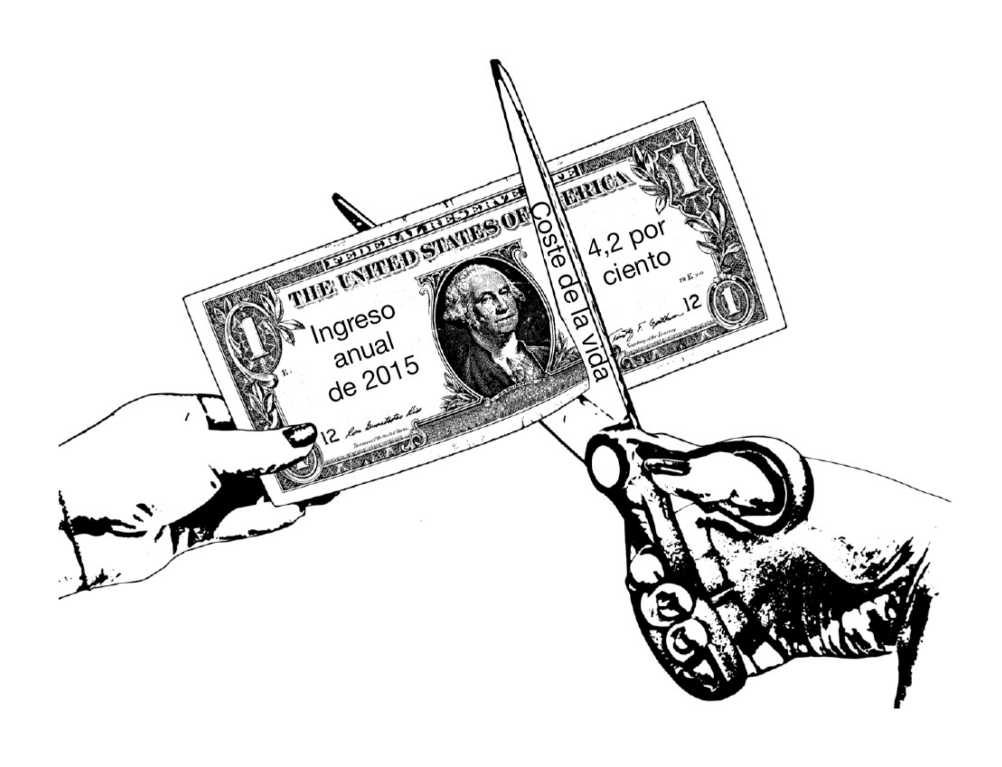
Las ventas de helados aumentan según se incrementa la cantidad de personas que llevan pantalones cortos. Ninguna de ambas cosas causa la otra; el tercer factor *x* que causa ambas son las temperaturas elevadas del verano. La cantidad de programas de televisión que aparecieron cada año cuando éramos niños podría estar correlacionada con nuestra estatura, pero no cabe duda de que lo que provocaba ambos eventos era el paso del tiempo en una época en la que (a) la televisión expandía su mercado y (b) estábamos creciendo.

¿Cómo sabemos cuándo una correlación indica una causa? Un modo de saberlo es realizar un experimento controlado. Otro es aplicar la lógica. Pero es necesario ser cauteloso, es fácil liarse con la semántica. ¿Fue la lluvia lo que causó que las personas se pusieran gabardinas o fue su deseo de no mojarse, como consecuencia de la lluvia, la causa?

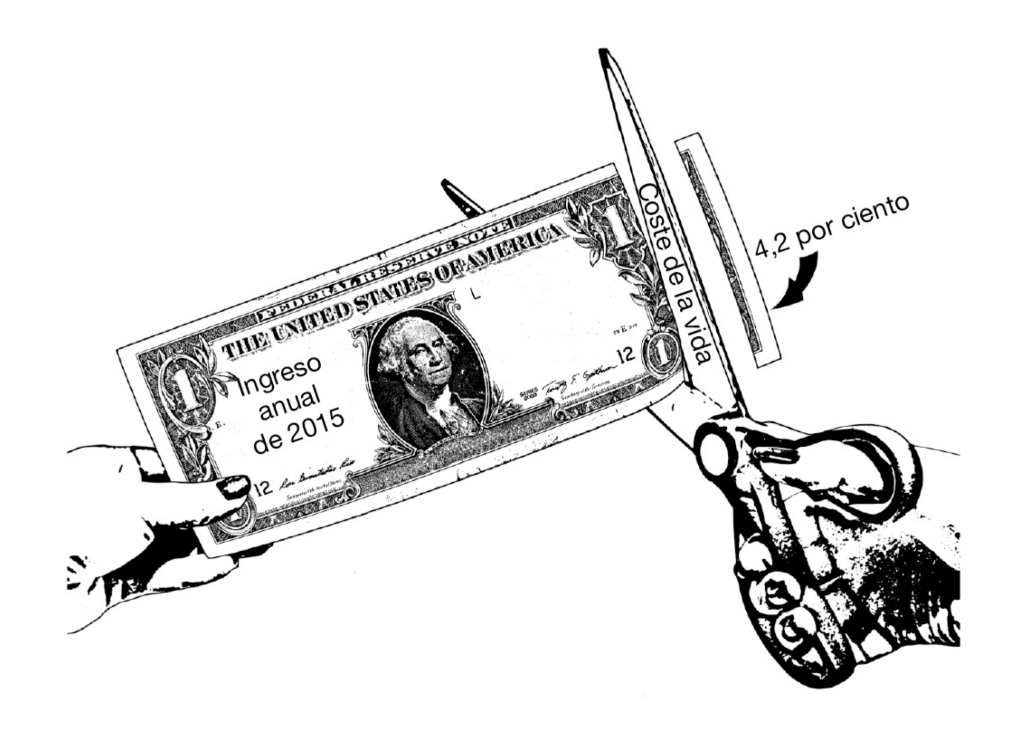
Esta idea ha sido ingeniosamente ilustrada por Randall Munroe[48] en su caricatura *xkcd* en Internet. Dos muñecos de palitos, en apariencia estudiantes universitarios, están hablando. Uno de ellos afirma que solía pensar que la correlación supone causación. Después fue a una clase de estadística y ahora ya no cree tal cosa. El otro le responde «Parece que la clase te ayudó». El primero contesta «Bueno, tal vez».

Ilustraciones engañosas

Las infografías a veces se emplean como artimañas para formar la opinión pública, basándose en el hecho de que la mayoría de las personas no las analizan con cuidado. Consideremos la siguiente ilustración, que podría utilizarse para atemorizarnos, haciéndonos creer que la inflación creciente está devorando nuestros ahorros:



Es una imagen aterradora, pero observemos con cuidado. Las tijeras no están cortando el billete un 4,2% de su tamaño. Cuando nuestro sistema visual se confronta con nuestro sistema lógico[49], habitualmente gana el sistema visual, a menos que nos esforcemos con ahínco para superar el sesgo visual. El infográfico correcto habría debido ser como sigue, pero habría tenido menos impacto emocional:



Interpretación y contexto

A menudo hay estadísticas que se elaboran y se muestran correctamente, pero alguien, un periodista, un abogado, cualquiera sin formación estadística, las representa erróneamente, bien porque no las ha comprendido o bien porque no se ha dado cuenta de que un cambio menor en la descripción puede alterar el significado.

A menudo quienes desean utilizar estadísticas no disponen de estadísticos en su plantilla, de manera que buscan las respuestas a sus preguntas en personas que carecen de la formación adecuada. Las grandes empresas, los departamentos del Gobierno, las organizaciones sin ánimo de lucro y las fruterías de la esquina se benefician de estadísticas sobre asuntos como las

ventas, los clientes, las tendencias y la cadena de abastecimiento. La incompetencia puede interferir en cualquier etapa del proceso, en el diseño experimental, en la recogida de datos, en su análisis e interpretación.

A veces la estadística sobre la que se informa no es la pertinente. Si estamos intentando convencer a unos accionistas de que nuestra empresa marcha bien, podríamos publicar una estadística sobre las ventas anuales, mostrando cifras que se incrementan gradualmente. Pero si el mercado del producto que fabricamos se está expandiendo, es de esperar que se produzca un incremento en las ventas. Lo que los inversores y analistas querrían saber es si nuestra cuota de mercado ha cambiado. Si nuestra cuota de mercado está disminuyendo porque la competencia se está llevando nuestros clientes, ¿qué podríamos hacer para que nuestro informe resultara atractivo? Es sencillo: no informaremos sobre la estadística pertinente de la cuota de mercado y nos limitaremos a informar sobre los datos de ventas. ¡Las ventas están aumentando! ¡Todo marcha bien!

Los perfiles financieros que aparecían en las solicitudes de hipoteca de hace veinticinco años probablemente no serían de gran ayuda para elaborar un modelo de riesgos actual. Cualquier modelo de conducta del consumidor en una web[50] puede quedar obsoleto muy rápidamente. Las estadísticas sobre la solidez del hormigón que se utiliza para construir pasarelas pueden no ser pertinentes para el hormigón que se emplea en los puentes (donde la humedad y otros factores podrían provocar una divergencia, incluso si ambos tipos de obra empleasen el mismo tipo de hormigón).

Probablemente conoceréis alguna de las variantes del anuncio «cuatro de cada cinco dentistas recomiendan Colgate». Es verdad. Lo que la agencia de publicidad que hay tras esos anuncios de hace décadas quiere que penséis es que los dentistas prefieren Colgate más que cualquier otra marca. Pero eso no es verdad. La Agencia de Control de la Publicidad del Reino Unido investigó ese anuncio y determinó que constituía una práctica desleal, porque la investigación que realizaron permitía a los dentistas recomendar más de una marca. De hecho, el

principal competidor de Colgate fue mencionado casi tan frecuentemente como Colgate[51] (un detalle que no veréis en los anuncios de Colgate).

Aludí al contexto en la sección sobre los promedios e implícitamente también al comentar los gráficos. La manipulación del contexto de cualquier mensaje habilita una enorme cantidad de posibilidades para que nos hagan creer algo que no es real, si no nos detenemos a considerar lo que nos están contando. La red de televisión por cable C-SPAN anuncia que está «disponible»[52] en 100 millones de hogares; lo que no significa que 100 millones de personas estén viendo C-SPAN, ni siquiera significa que una sola persona lo esté haciendo[53].

Las manipulaciones del contexto pueden influir sobre las políticas públicas. Una investigación sobre el reciclaje de basuras en varias calles de Los Ángeles muestra que una calle en concreto recicla 2,2 veces más que cualquier otra. Antes de que el ayuntamiento conceda a los residentes de esa calle un premio por sus esfuerzos ecológicos, preguntémonos a qué puede deberse tal resultado Una posibilidad es que esa calle tenga más del doble de residentes que las demás, quizá porque es más larga o tal vez porque hay más edificios de apartamentos. Medir el reciclado de la calle no es la estadística pertinente a menos que todas las demás calles sean idénticas en todo lo restante. Una estadística mejor sería calcular por unidad familiar (medir lo que recicla cada familia) o, mejor aún, ya que las familias numerosas probablemente consumen más que las familias pequeñas, se podría calcular por individuo. Es decir, es necesario ajustar la cantidad de basura reciclada recogida en función de la cantidad de personas que viven en cada calle. Ese es el *verdadero contexto* de la estadística.

El periódico *Los Angeles Times* informaba en 2014 sobre el uso del agua en la ciudad de Rancho de Santa Fe[54], en la California agobiada por la sequía. «En términos diarios per cápita, en septiembre los hogares de esa región absorbieron un promedio de casi cinco veces más agua que los hogares de la costa del sur de California, obteniendo el dudoso galardón de ser

los mayores acaparadores residenciales de agua del Estado». «Hogares» no es el contexto pertinente para esta estadística y *Los Angeles Times* estaba en lo cierto al informar en términos per cápita, por individuo; tal vez los residentes de Rancho de Santa Fe tienen familias más numerosas, lo que implica más duchas, más platos y más descargas de la cisterna. Otro contexto hubiera podido considerar el uso de agua por acre[55]. Las casas de Rancho de Santa Fe tienden a tener parcelas grandes; tal vez resulte deseable para la prevención de los incendios, y por otros motivos, mantener las tierras plantadas con vegetación y las grandes parcelas de Rancho de Santa Fe no utilizan más agua por acre que cualquier otro lugar.

De hecho, hay una indicación en ese sentido en un artículo del *New York Times* sobre este asunto: «Los responsables públicos del agua han advertido contra la comparación del uso de agua per cápita entre distritos; afirman que se espera que el gasto sea mayor en las comunidades acaudaladas con parcelas grandes».

El problema con los artículos de periódico es que dotan de un contexto a los datos que lleva a pensar que los residentes de Rancho de Santa Fe están utilizando más agua de la que les corresponde, pero los datos que proporcionan, como sucede en el caso del reciclado de Los Ángeles citado anteriormente, no muestran tal cosa.

Calcular las proporciones en lugar de las cifras brutas frecuentemente ayuda a establecer el contexto correcto. Supongamos que eres el representante comercial para el noroeste de una empresa que vende capacitadores de flujo[56]. Tus ventas han aumentado mucho, pero todavía no puedes competir con tu pesadilla en la empresa, Jack, el representante para el suroeste. Es poco justo, su territorio no solo es mayor, sino que comprende mucha más población. Los bonus en tu empresa dependen de que demostréis tener el empuje necesario para echaros a la calle y conseguir ventas.

Hay un modo legítimo de presentar tu caso: informar de las ventas en función de la extensión o de la población del territorio que atendéis. En otras palabras, en lugar de mostrar en un gráfico la cantidad total de capacitadores de flujo vendidos,

exponer la cifra por persona en esa región o por kilómetro cuadrado. En ambos casos saldrás beneficiado.

Informes recientes han mostrado que 2014 fue uno de los peores años en accidentes de aviación: 22 accidentes produjeron 992 muertos. Pero resulta que volar es hoy día más seguro que nunca[57]. Dado que en la actualidad hay muchos más vuelos que antes, 992 muertes representan un descenso espectacular en la cantidad de muertes por millón de pasajeros (o por millón de millas de vuelo). En cualquier vuelo de una aerolínea importante, las probabilidades de matarse son de 1 en 5 millones, lo que significa que es más probable matarse haciendo prácticamente cualquier otra cosa, cruzando la calle, comiendo (la muerte por ahogamiento o por envenenamiento involuntario es unas 1.000 veces más probable). En este caso, la línea de base de la comparación es muy importante. Estas estadísticas abarcan un año, un año de viajes en avión, un año de comer y asfixiarse o envenenarse. Si cambiáramos la línea de base y considerásemos cada hora de cada una de estas actividades, la estadística también cambiaría.

Diferencias que no suponen una diferencia

A menudo se utilizan estadísticas para intentar entender si existe una diferencia entre dos tratamientos: el uso de dos fertilizantes diferentes, dos medicamentos contra el dolor, dos estilos de enseñanza, dos grupos salariales (por ej., hombres y mujeres que hacen el mismo trabajo). Dos tratamientos pueden diferir de muchas maneras; puede haber diferencias reales entre ellos; puede haber factores de distorsión en la muestra que no guarden relación con los tratamientos; puede haber errores de medida; o pueden producirse variaciones al azar, pequeñas diferencias que surgen por azar, a veces a un lado de la ecuación, a veces al otro, dependiendo de cuando se observe. El objetivo del investigador es buscar diferencias estables y replicables y lo que pretendemos es diferenciarlas del error experimental.

Sin embargo, hay que estar sobre aviso del empleo de la palabra «significativo» en los medios de comunicación, ya que para los estadísticos eso no significa «notable». En estadística, la palabra «significativo» quiere decir que los resultados han superado pruebas matemáticas como la prueba de *t,* la de ji-cuadrado, de regresión y el análisis de componentes principales (hay cientos). Las pruebas de significación estadística cuantifican hasta qué punto el mero azar puede explicar unos resultados. Cuando se dispone de una gran cantidad de observaciones, incluso diferencias pequeñas, de una dimensión de magnitud trivial, pueden situarse fuera de lo que puede explicarse mediante nuestros modelos de aleatorización y cambio. Esas pruebas no nos dicen lo que es notable y lo que no lo es, eso es el ámbito del juicio humano.

Cuantas más observaciones tengamos de los dos grupos, más probable resultará que encontremos diferencias entre ellos. Supongamos que pongo a prueba los costes de mantenimiento anuales de dos automóviles, un Ford y un Toyota, registrando los gastos en reparaciones de diez vehículos de cada marca. Digamos que hipotéticamente el coste medio de mantener el Ford es de ocho céntimos más por año. Probablemente esa cifra no alcanzaría el nivel de significación estadística y es evidente que una diferencia de ocho céntimos al año no se va a convertir en un factor decisorio sobre qué coche comprar, es demasiado pequeña para preocuparse. Pero si observásemos los gastos de reparaciones de 500.000 vehículos, esa diferencia de ocho céntimos sería estadísticamente significativa. Sin embargo, es una diferencia que no tiene relevancia en el mundo real, en un sentido práctico. De modo semejante, un fármaco nuevo contra el dolor de cabeza puede que lo cure más rápidamente, pero ¿a quién le importa si es solo 2,5 segundos más rápido?

Interpolación y extrapolación

Salís al jardín y observáis una planta de diente de león que el martes tiene una altura de ocho centímetros. Volvéis a mirar el

jueves y tiene doce centímetros. ¿Qué altura tenía el miércoles? No lo sabemos con seguridad porque el miércoles no la medimos (el miércoles es el día que os quedasteis atascados en el tráfico volviendo a casa desde la guardería, cuando comprasteis el herbicida). Podemos intentar adivinar: es probable que el diente de león midiera diez centímetros el miércoles. Esto es una interpolación. En las interpolaciones se consideran dos puntos de observación de datos y se hace una estimación de lo que habría ocurrido entre ellos si se hubiera tomado una medida.

¿Cuán alto será el diente de león dentro de seis meses? Si crece a un ritmo de dos centímetros por día, podríamos decir que crecerá 360 centímetros en seis meses (aproximadamente 180 días), hasta alcanzar un total de 372 centímetros, o tres metros y setenta y dos centímetros. Ahora estáis empleando una extrapolación. Pero ¿habéis visto alguna vez un diente de león de ese tamaño? Lo más seguro es que no. Se doblan por su propio peso o mueren por causas naturales, o alguien los pisa, o los mata el herbicida. La interpolación no es una técnica perfecta, pero si las dos observaciones que estamos considerando están muy próximas, normalmente la interpolación nos proporciona una buena estimación. Sin embargo, la extrapolación es más arriesgada, ya que estamos realizando estimaciones que se sitúan fuera del rango de nuestras observaciones.

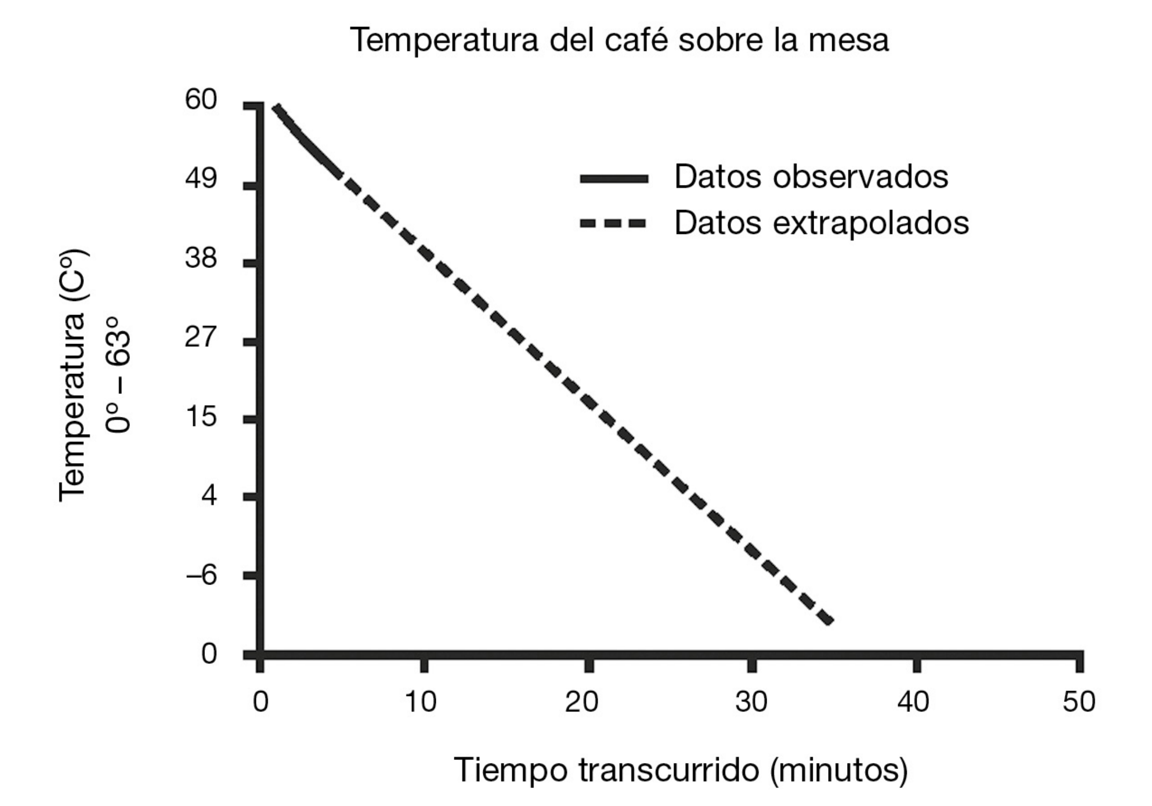
La cantidad de tiempo que una taza de café tarda en alcanzar la temperatura ambiente de una habitación se rige por la ley del enfriamiento de Newton[58] (y es afectada por otros factores como la presión barométrica y la composición de la taza). Si el café empezó a 63°, podríais observar que la temperatura disminuye del siguiente modo:

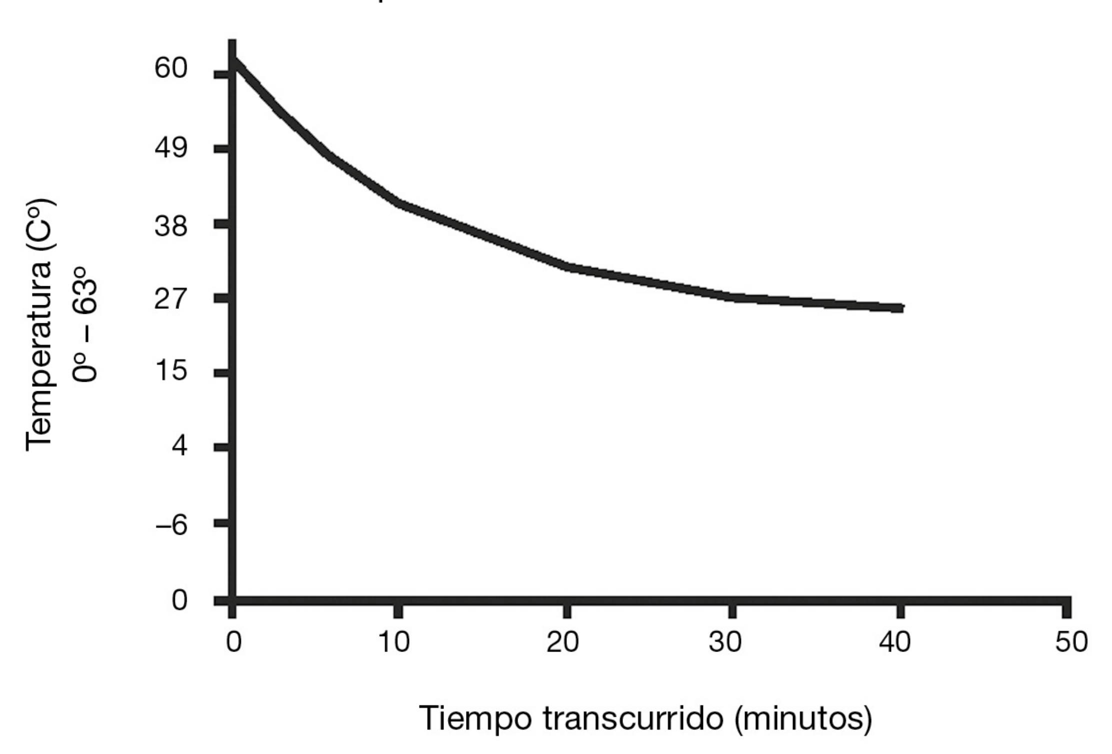
Tiempo transcurrido (mim.)	*Temperatura*
0	62,7°
1	60,0°
2	57,2°

3	54,4°

El café pierde aproximadamente 2,7° cada minuto. Si interpoláis entre dos observaciones, o sea, si quisierais conocer la temperatura que tendría entre dos puntos de medida, la interpolación sería bastante precisa. Pero si extrapoláis a partir de este patrón de datos, lo más probable es que lleguéis a una respuesta absurda, como que el café alcanzará el punto de congelación una vez que hayan transcurrido 30 minutos (véase el gráfico de la página siguiente).

La extrapolación no llega a tomar en cuenta una limitación física: el café no puede enfriarse más que la temperatura de la habitación. Tampoco tiene en cuenta que la tasa de enfriamiento del café se hace más lenta a medida que se aproxima a la temperatura de la habitación. El resto de la función de enfriamiento tiene esta apariencia:





Obsérvese que la inclinación de la curva durante los diez primeros minutos no se mantiene, se va aplanando. Esto permite resaltar la importancia de dos asuntos cuando extrapolamos: disponer de una cantidad grande de observaciones que abarque un rango amplio y disponer de algún conocimiento sobre el proceso subyacente.

Precisión frente a exactitud

Cuando valoramos la precisión de las cifras tendemos a creer que también son *exactas,* pero no es lo mismo. Si afirmo que «muchas personas compran coches eléctricos en la actualidad», parece que estoy haciendo una suposición. Pero si digo «el 16,39% de las ventas de vehículos nuevos son coches eléctricos», damos por supuesto que sé de lo que estoy hablando. En este caso estaríamos confundiendo precisión con exactitud. Podría habérmelo inventado. Podría haber realizado un muestreo de una cantidad pequeña de personas que viven cerca de un concesionario de vehículos eléctricos.

Recordemos el titular de la revista *Time* que mencioné anteriormente, en el que se afirmaba que hay más personas con teléfonos móviles que con W.C. No resulta inverosímil, pero se trata de una distorsión, ya que esto no es lo que descubrió el estudio de la ONU. Lo que la ONU indicó es que había más personas con *acceso* a teléfonos móviles que a cuartos de baño, lo que, como sabemos, es un asunto diferente. Un teléfono móvil puede ser compartido por docenas de personas. La falta de servicios de aseo sigue siendo preocupante, pero el titular sugiere que si los contáramos encontraríamos más teléfonos móviles que cuartos de baño en el mundo, y eso no es lo que indican los datos.

Acceso es una de esas palabras que debiera ser señalada con una bandera roja cuando nos la encontramos en una estadística. Que las personas dispongan de acceso a servicios de salud puede significar simplemente que viven cerca de un hospital, no que serían admitidas en este o que pudieran costeárselo. Como vimos anteriormente, C-SPAN es accesible en 100 millones de hogares[59], pero eso no significa que lo vean 100 millones de personas. Podría afirmar que el 90% de la población mundial tiene acceso a *La mentira como arma,* mostrando que el 90% de la población vive a menos de 40 kilómetros de una conexión a Internet[60], de un ferrocarril, de una carretera, de una pista de aterrizaje, de un puerto o de una ruta de trineo.

Comparar peras con manzanas

Uno de los modos de mentir con la estadística es comparar conjuntos de datos, poblaciones, tipos de productos, que son diferentes entre sí y pretender que no lo son. Como dice el refrán, no se pueden comparar peras con manzanas.

Haciendo uso de métodos dudosos, alguien podría afirmar que resulta más seguro ser militar durante un conflicto activo (como en la guerra actual en Afganistán) que quedarse en los Estados Unidos, cómodamente instalado en casa. Comencemos con los 3.482 militares estadounidenses en servicio activo que murieron

en 2010[61]. De un total de 1.431.000 militares[62], esto supone una tasa de 2,4 muertos por cada 1.000 personas. En el conjunto de los Estados Unidos la tasa de mortalidad fue de 8,2 muertes por cada 1.000 personas[63]. En otras palabras, es tres veces más seguro estar en el ejército, en una zona de guerra, que vivir en los Estados Unidos.

¿Cuál es el problema? Que las dos muestras no son semejantes y no se pueden comparar directamente. El personal militar en activo tiende a ser joven y a gozar de buena salud; se alimentan con una dieta nutritiva y tienen un buen servicio de salud. La población general de los Estados Unidos[64] incluye a los ancianos, personas enfermas, miembros de bandas juveniles, adictos al crack, conductores temerarios, lanzadores de puñales y a muchas personas que ni disponen de una dieta nutritiva, ni disfrutan de buenos servicios de salud; su mortalidad sería elevada dondequiera que estuvieran. Por su parte, no todos los militares están localizados en zonas de guerra, algunos de ellos residen en bases completamente seguras en los Estados Unidos, otros están sentados en sus oficinas del Pentágono o trabajan en los centros de reclutamiento en los centros comerciales de los barrios periféricos.

La revista *US News & World Report* publicó un artículo comparando la proporción de republicanos y demócratas en el país desde los años treinta. El problema es que los métodos de muestreo han cambiado a lo largo del tiempo. En los años treinta y cuarenta los muestreos habitualmente se hacían mediante entrevistas individuales y listas de correo generadas a partir del listín telefónico; hacia los años setenta el muestreo era básicamente telefónico. El muestreo en la primera mitad del siglo XX estaba sesgado a favor de quienes tuvieran teléfono fijo: las personas más adineradas, quienes, al menos en esos tiempos, tendían a votar republicano. Hacia el año 2000, se empleaban los móviles para el muestreo, lo que lo sesgó hacia los jóvenes, quienes tendían a votar demócrata. No es posible saber si la proporción de demócratas y republicanos realmente ha cambiado desde los años treinta, ya que las muestras no son equiparables.

Creemos que estamos estudiando una cosa, pero en realidad estudiamos otra.

Un problema semejante tiene lugar cuando se informa sobre el descenso de las muertes por accidentes de tráfico de moto en la actualidad en comparación con hace tres décadas. Los datos más recientes pueden incluir las motos de tres ruedas, comparadas con las motos de dos ruedas del último siglo; puede que se compare un periodo en el que el uso del casco no era obligatorio con la actualidad, en que lo es en la mayoría de los estados.

¡Prestad atención a los cambios de las muestras antes de sacar conclusiones! La revista *US News & World Report* (sí, ellos otra vez) informó de un incremento del número de médicos[65] en un periodo de doce años, acompañado de un descenso significativo del salario medio. ¿Cuál es el mensaje a recordar? Podría concluirse que no es un buen momento para ingresar en la profesión médica, ya que hay una avalancha de doctores y el exceso de oferta que supera a la demanda ha disminuido el sueldo de los médicos. Puede que sea verdad, pero no hay pruebas que apoyen esa afirmación.

Un argumento igualmente verosímil es que en un periodo de doce años el aumento de la especialización y el crecimiento de la tecnología han creado más oportunidades para los médicos, de modo que hay más trabajos en oferta, lo que aumenta la cantidad total de médicos. ¿Y el descenso del salario? Quizá muchos de los médicos de más edad se han jubilado y han sido sustituidos por otros más jóvenes, que ganan sueldos inferiores al estar recién licenciados. No hay pruebas de una cosa o de la otra. Una parte importante de la alfabetización estadística consiste en darse cuenta de que algunas estadísticas no pueden ser interpretadas tal y como han sido presentadas.

A veces estas comparaciones de peras con manzanas dan lugar a submuestras inconsistentes, al ignorar un detalle cuya importancia no percibimos. Por ejemplo, cuando se obtiene una muestra de maíz de un campo que se ha abonado con un fertilizante nuevo, podríamos no tomar nota de que algunas mazorcas de maíz reciben más sol y algunas reciben más agua. O cuando estudiamos cómo actúan los patrones de la circulación

sobre la repavimentación de las calles, podríamos no darnos cuenta de que algunas calles tienen más vertidos de agua que otras, lo que influye sobre la necesidad de renovar el asfalto.

Amalgamar es reunir cosas que son diferentes (heterogéneas) en el mismo bote o categoría, una forma de mezclar peras con manzanas. Si analizamos la cantidad de piñones de engranaje defectuosos producidos por una fábrica, podríamos combinar dos tipos completamente diferentes para conseguir que las cifras resultaran más favorables a nuestros intereses particulares.

Veamos un ejemplo de la política pública. Podríamos estar interesados en supervisar la conducta sexual de los adolescentes y preadolescentes. Cómo amalgamamos (o agrupamos) los datos puede tener un efecto importante sobre cómo las personas perciben nuestros datos. Si lo que pretendemos es obtener financiación para los centros educativos y de orientación, qué mejor que aportar una estadística que diga «el 70% de los niños en edad escolar entre los diez y los dieciocho años son sexualmente activos». No nos sorprende que los de diecisiete y dieciocho años lo sean, pero ¡los de diez! Ese sondeo seguramente provocaría que los abuelos buscaran las sales aromáticas y empezaran a rellenar cheques. Sin embargo, resulta evidente que una única categoría formada desde los diez hasta los dieciocho años agrupa juntos individuos que probablemente sean sexualmente activos con otros que no lo son. Es más claro separar las latas y poner juntos a los individuos de edad semejante y experiencias parecidas: de los diez a los once, de los doce a los trece, de los catorce a los quince y de los dieciséis a los dieciocho, por ejemplo.

El anterior no es el único problema. ¿Qué significa «sexualmente activo»? ¿Cuál es la pregunta que se formuló a los escolares? ¿Se preguntó algo a los escolares? Tal vez se preguntó a sus padres. En esa cifra pueden incorporarse todo tipo de sesgos. Queda abierto a la interpretación el significado de «sexualmente activo». Las respuestas pueden variar mucho dependiendo de cómo se defina y, desde luego, puede que quienes respondan no estén diciendo la verdad (sesgo del informante).

Otro ejemplo: cabe que nos refiramos al desempleo como un problema general, pero esto supone el riesgo de mezclar personas y factores desencadenantes de orígenes muy diferentes. Algunas son personas con discapacidad y no pueden trabajar; otros han sido despedidos por motivos justificados, porque robaban o se emborrachaban en el trabajo; algunos quieren trabajar, pero carecen de la formación necesaria; otros están en la cárcel; algunos ya no quieren trabajar porque asisten a centros de formación, o han ingresado en un monasterio, o viven de las rentas de su familia. Cuando las estadísticas se utilizan para influir sobre las políticas públicas, o para recabar donaciones para una causa, o para hacer el titular de un periódico, a menudo quedan fuera los matices y éstos pueden suponer una gran diferencia.

Los matices con frecuencia nos cuentan una historia[66] sobre el patrón de los datos. No todas las personas acaban en el desempleo por las mismas razones. La probabilidad de que un alcohólico o un ladrón sean desempleados puede ser cuatro veces mayor que la de alguien que no lo sea. Estos patrones implican información que se pierde al amalgamar. Incluir estos factores entre los datos nos ayuda a observar quién está desempleado y por qué, puede conducir a mejores programas de formación para las personas que los necesitan o más centros de Alcohólicos Anónimos en una ciudad que no cuenta con los necesarios.

Si las personas y las instituciones que supervisan la conducta utilizan definiciones diferentes de las cosas, o procedimientos diferentes para medirlas, los datos que se incluyen en las estadísticas pueden ser muy disímiles o heterogéneos. Si intentamos determinar la cantidad de parejas que conviven pero no están casados, podríamos apoyarnos en datos previamente recopilados por distintas instituciones locales y estatales. Sin embargo, las definiciones divergentes pueden dar lugar a un problema de categorización: ¿qué es convivir? ¿Se determina por la cantidad de noches a la semana que duermen juntos? ¿Por el lugar donde tienen sus propiedades? ¿Donde reciben el correo? Algunas jurisdicciones reconocen las parejas del mismo sexo,

otras no. Si utilizamos datos de orígenes diversos, en los que se han aplicado estrategias diferentes, la estadística resultante podría tener muy poco significado. Si las prácticas de registro, compilación y medida divergen demasiado entre los distintos puntos de recogida de datos, la estadística podría no significar lo que creemos que significa.

Un informe reciente descubrió que la tasa de desempleo juvenil en España es de un alucinante 60%. El informe amalgamaba en la misma categoría personas que normalmente aparecen en categorías separadas: se contabilizó como desempleados a estudiantes que no buscan empleo, junto con trabajadores que acababan de ser despedidos y trabajadores que buscan empleo.

En los Estados Unidos hay *seis* índices diferentes (que van desde el U1 hasta el U6) para registrar el desempleo[67] (tal como se mide por el Departamento de Estadísticas Laborales), que reflejan diferentes interpretaciones de lo que significa realmente «desempleado». Pueden incluir personas que buscan un trabajo, personas que están en formación pero no buscan empleo, personas que buscan un empleo a tiempo completo en una empresa para la que ya trabajan a tiempo parcial, etc.

El periódico *USA Today* informó en julio de 2015 que la tasa de paro había caído[68] al 5,3%, «su nivel más bajo desde abril de 2008». Algunas fuentes más minuciosas, como *AP, Forbes* y el *New York Times,* daban cuenta de la razón de la caída[69]: muchas personas que estaban desempleadas habían dejado de buscar trabajo, por lo que técnicamente habrían quedado fuera de la mano de obra.

Amalgamar no siempre es incorrecto. Podemos optar por combinar las notas de los chicos y las chicas de una escuela, en especial si no hay pruebas de que sus puntuaciones difieran, de hecho, es una buena idea hacerlo a fin de aumentar el tamaño de la muestra (lo que proporciona una estimación más estable de lo que estamos estudiando). La definición excesivamente general de una categoría (como la del informe sobre la actividad sexual que mencionamos anteriormente) o las definiciones inconsistentes

(como la de la estadística sobre las parejas que cohabitan) presentan problemas de interpretación. Cuando se realiza adecuadamente, la amalgama nos ayuda a obtener un análisis válido de los datos.

Supongamos que trabajáis para el estado de Utah y que un gran fabricante nacional de ropa para bebés está planteándose trasladarse a vuestro Estado. Pensáis que si sois capaces de mostrar que en Utah hay muchos nacimientos, estaréis en mejor posición para atraer a esa empresa, así que vais a la web Census.gov y sacáis un gráfico de la cifra de nacimientos por Estado:

Nacimientos: Estados Unidos, 2013

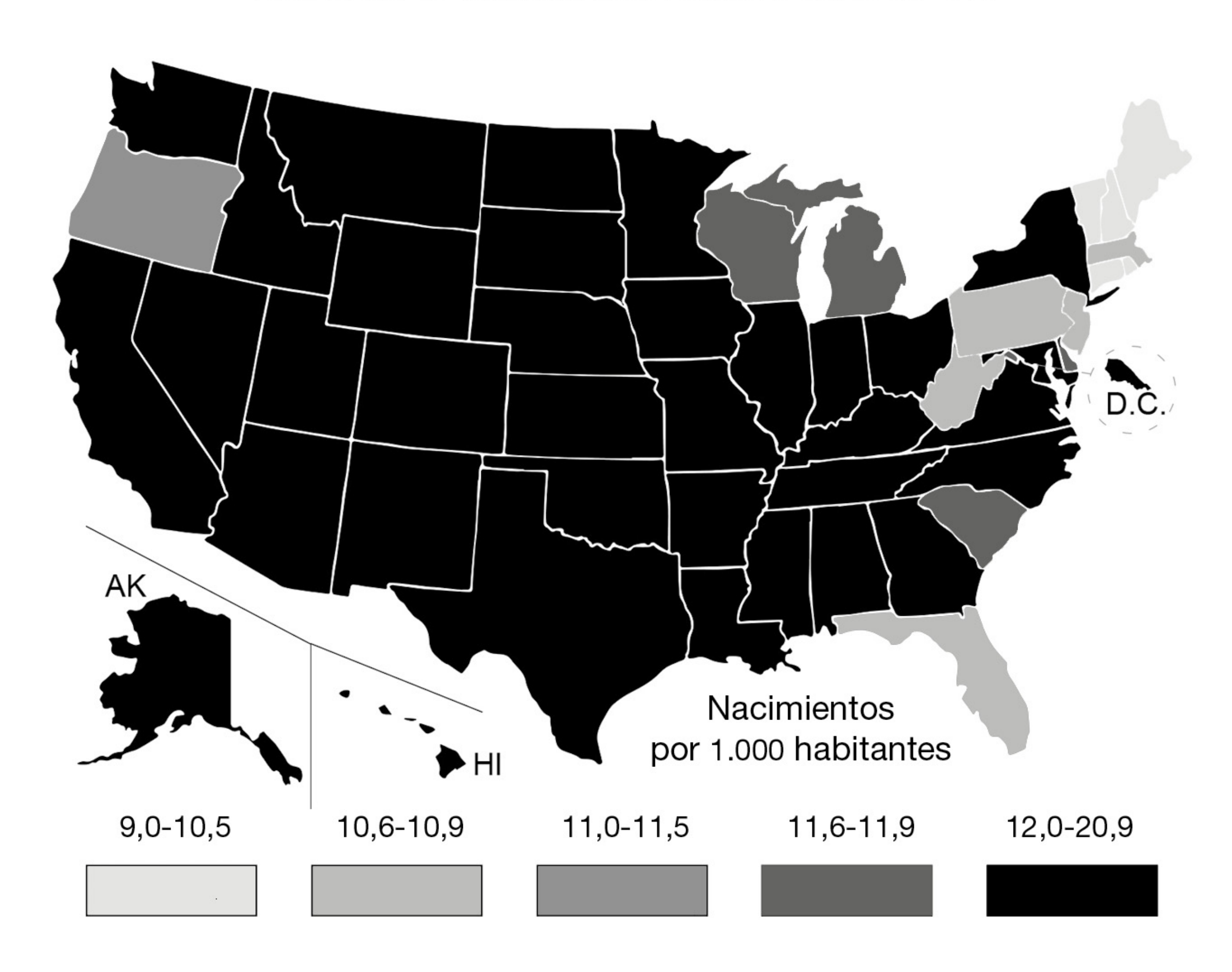
Utah
D.C.
AK
HI

Nacimientos
por 1.000 habitantes

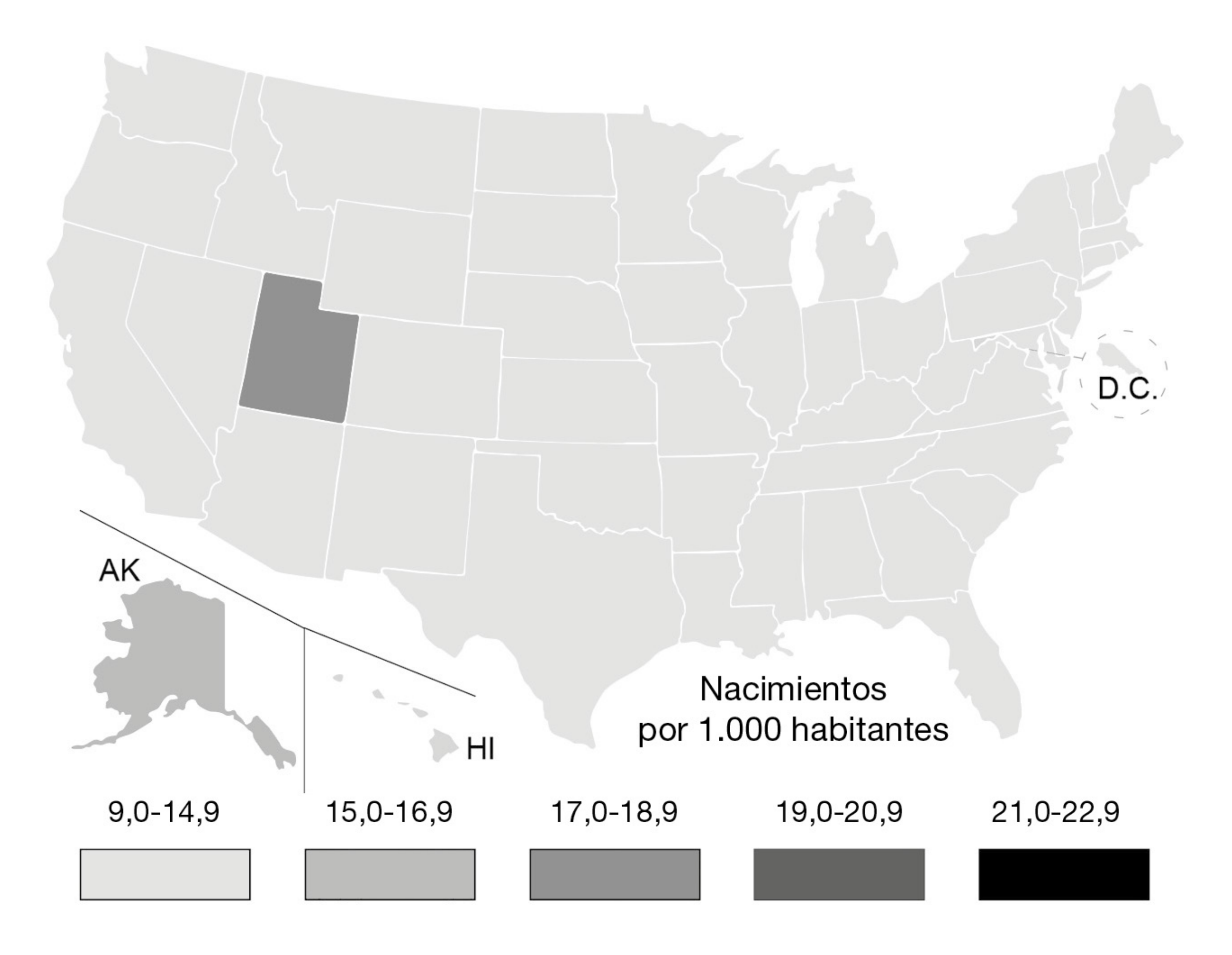
5,9-13,9
14,0-39,9
40,0-68,9
69,0-149,9
150,0-494,8

Utah parece mejor que Alaska, el Distrito Federal, Montana, Wyoming, las Dakotas y los estados pequeños del noreste. Pero no es ninguna explosión demográfica si se compara con California, Texas, Florida y Nueva York. Pero, un momento, el mapa que hemos obtenido muestra la *cifra bruta* de nacimientos,

así que está seriamente sesgado a favor de los Estados con mayor población. Podríamos hacer un gráfico con la *tasa* de nacimientos por mil habitantes de la población:

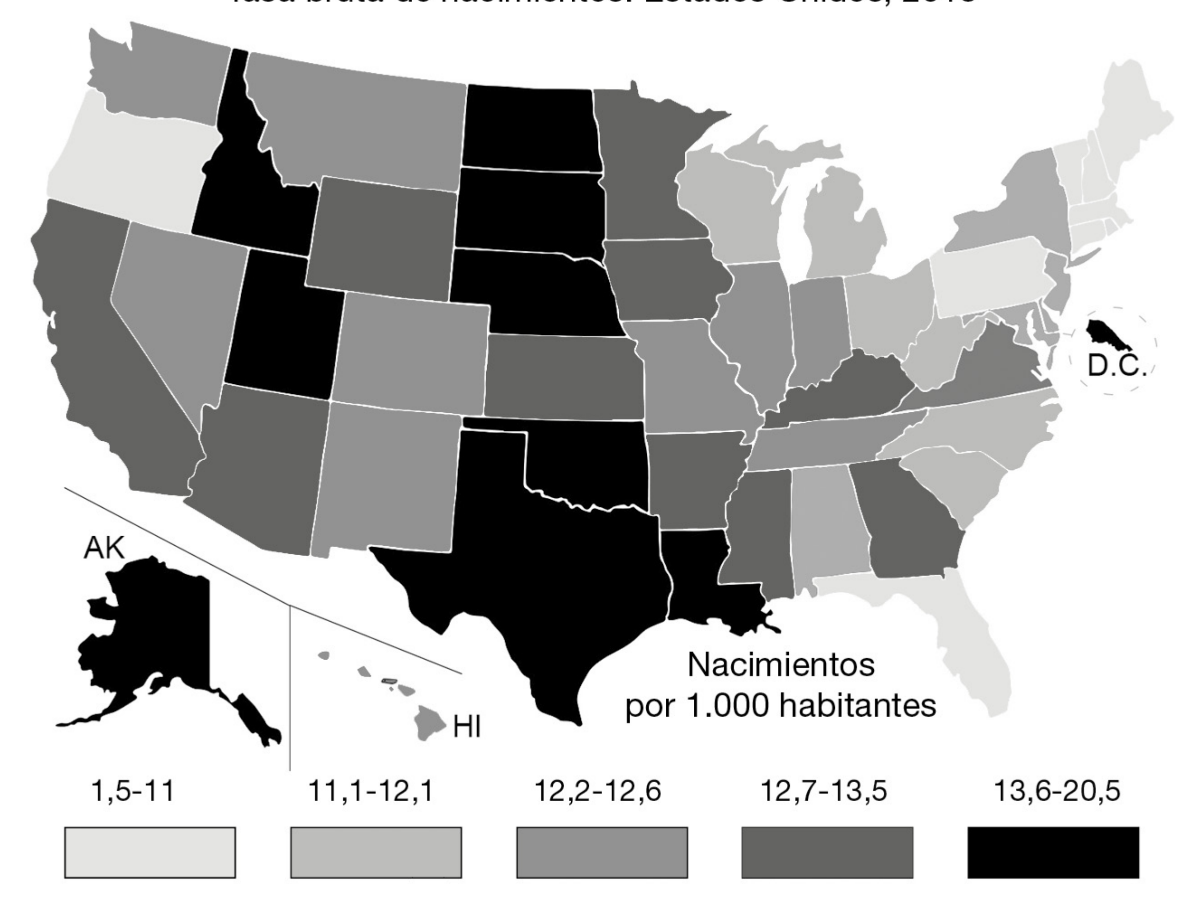


Esto no nos ayuda, Utah aparece como la mayor parte del país. ¿Qué hacemos? ¡Cambiamos las categorías! Podemos cacharrear con el rango de valores en cada categoría, esos que tienen cinco tonos del gris al negro en la parte de abajo. Si nos aseguramos de que Utah se sitúe en una categoría exclusiva de ella, podemos conseguir que destaque sobre el resto del país.

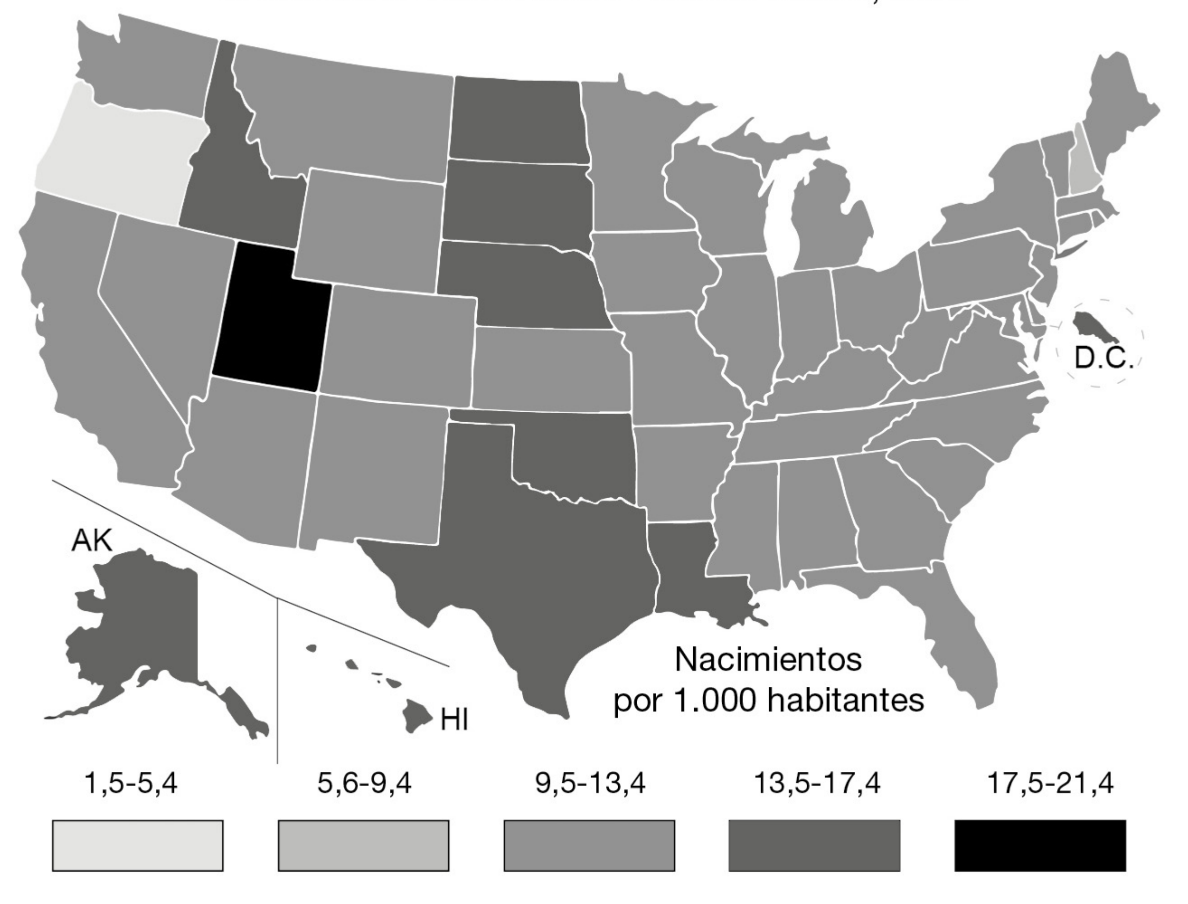


Está claro que esto únicamente funciona porque en realidad Utah tiene la tasa de nacimientos más elevada del país, no por mucha diferencia, pero es la más alta. Al situarlo en solitario, con un color diferente, hemos logrado que destaque. Si estuviéramos intentando defender la posición de otros estados, habríamos tenido que recurrir a otros juegos de manos, como hacer un gráfico de la cantidad de nacimientos por kilómetro cuadrado, por tienda Walmart o como función de los ingresos disponibles. Si maniobramos el tiempo suficiente, encontraremos alguna métrica para defender la posición de cualquiera de los cincuenta estados.

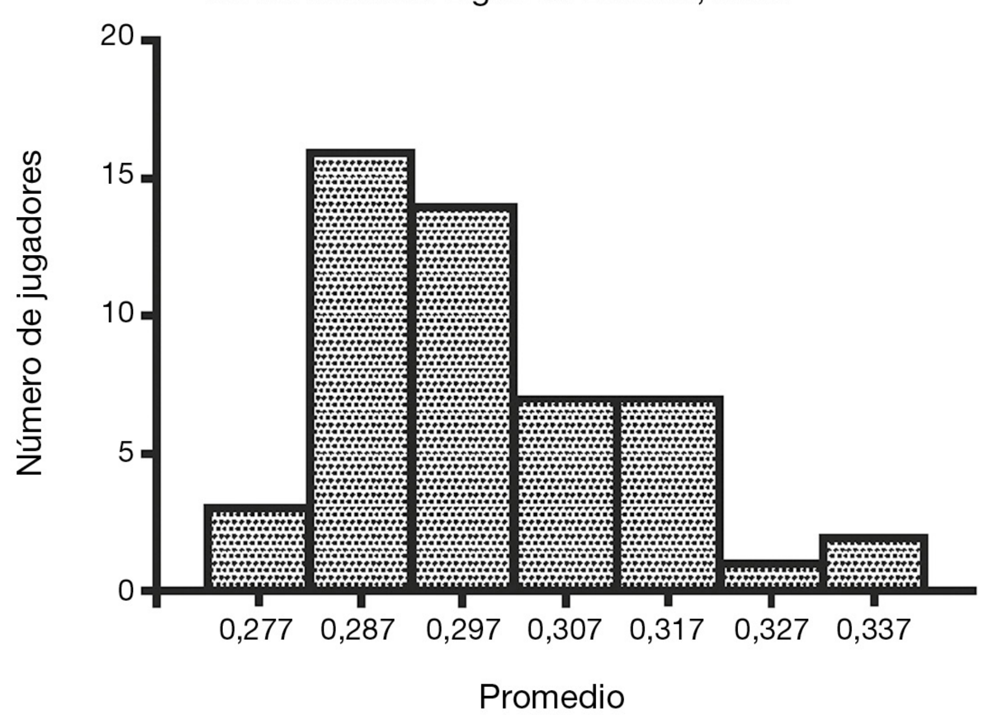
¿Es ese el modo *correcto,* el modo honesto de presentar un gráfico? Es una cuestión a valorar, pero un modo relativamente neutro sería agrupar los datos de modo que el 20% de los estados caiga en una de cinco categorías, es decir, una cantidad igual de estados en cada color:



Otro modo sería hacer que cada categoría tuviera el mismo tamaño:



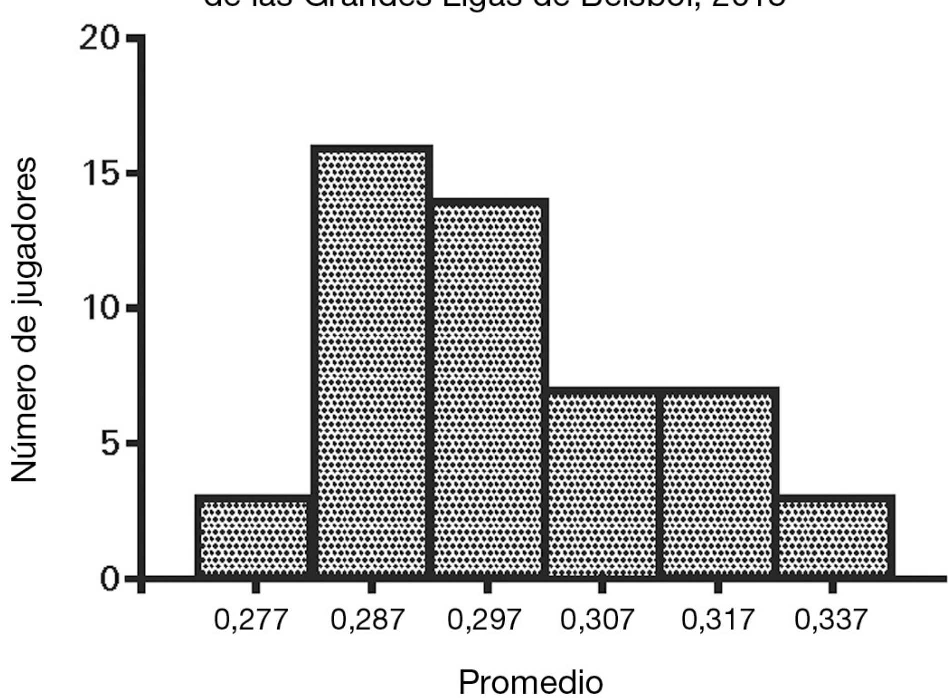
Este tipo de artimaña estadística, emplear categorías con anchos diferentes en todos menos en el último de los mapas, aparece con frecuencia en los histogramas, donde habitualmente las categorías se identifican mediante su punto medio, y hemos de inferir su rango por nuestra cuenta. Seguidamente aparecen los promedios de bateo de la temporada 2015 de los 50 mejores jugadores de las Grandes Ligas de Béisbol (National and American Leagues):



*Estadísticas tomadas de http://mlb.com/stats/sortable.jsp#elem=[object+Object]&tab_level=child&click_text=Sortable+Player+hitting&game–type=%27R&season=2015&season_type=ANY&league_code=%27MLB%27sectionType=sp&statType=hitting&page=1&ts=1457286793822&palyerType=QUALIFIER&timeframe=.

Bien, supongamos que somos el jugador cuyo promedio de bateo es 0,330, lo que nos sitúa en la segunda categoría más alta. Ha llegado el momento de reclamar nuestras primas y no tenemos la menor intención de ofrecer a la dirección del equipo un motivo para que nos nieguen la prima esta temporada, ya nos hemos comprado un Tesla. Así que cambiamos los anchos de las categorías, amalgamamos nuestros resultados con los de los dos jugadores que batean 0,337 y ahora nos encontramos en el grupo de los mejores jugadores. Cuando lo hagamos es importante hacer desaparecer el hueco que queda (ya no habrá jugadores en la categoría cuyo punto central es 0,327), para ello creamos una discontinuidad en el eje *x* que probablemente pocos percibirán:





Subdivisión engañosa

Lo opuesto de amalgamar es subdividir y esto puede conducir a la gente a creer todo tipo de falsedades. Para afirmar que *x* es la causa de *y* simplemente necesito subdividir otras causas en categorías cada vez menores.

Supongamos que trabajáis en una fábrica de purificadores de aire y organizáis una campaña para demostrar que las enfermedades respiratorias son la principal causa de muerte en los Estados Unidos, superando a otras causas como las enfermedades cardíacas y el cáncer. En la actualidad la verdadera causa principal de muerte en los Estados Unidos son las enfermedades cardíacas. El Centro de Control de Enfermedades de Estados Unidos informa que las tres principales causas de muerte en 2013 fueron las siguientes[70]:

Enfermedad cardíaca: 611.105

Cáncer: 584.881

Enfermedades respiratorias crónicas: 149.205

Pues bien, dejando al margen el inoportuno detalle de que los purificadores de aire domésticos pueden no ser una defensa relevante contra las enfermedades respiratorias crónicas, estas cifras no son demasiado favorables para vuestra empresa. Está claro que os gustaría salvar más de 100.000 vidas al año, pero afirmar que estáis combatiendo la *tercera* mayor causa de muerte no va suponer una campaña demasiado impresionante. Pero ¡un momento! La enfermedad cardíaca no es una sola cosa, sino varias:

Fiebre reumática aguda y enfermedad cardíaca reumática crónica: 3.260
Enfermedad cardíaca hipertensa: 37.144
Infarto agudo de miocardio: 116.793
Fallo cardíaco: 65.120

Esto marcha. Ahora vamos a descomponer el cáncer en subtipos menores. ¡Al no amalgamar y establecer estas refinadas subdivisiones, lo habéis logrado! Las enfermedades respiratorias crónicas se han convertido en la principal causa de muerte. Os acabáis de ganar un sobresueldo. Algunas empresas de alimentación han empleado esta estrategia de subdivisión para ocultar las cantidades de grasas y azúcares que contienen sus productos.

---

46 http://qz.com/122921/the-chart-tim-cook-doesnt-want-you-to-see/; http://www.tekrevue.com/tim-cook-trying-prove-meaningless-chart/.

47 http://www.tylervigen.com/spurious-correlations.

48 https://xkcd.com/552/.

49 Este ejemplo se basa en el de Huff, *ibid*.

50 Se trata de una cita prácticamente literal de R. D. De Veux y D. J. Hand (2005), «How to lie with bad data», *Statistical Science,* 20(3), 231-238, p. 232.

51 Agradezco a mi alumna Vivian Gu este ejemplo.
D. Derbyshire (17 de enero de 2007), «Colgate gets the brush off for “misleading” ads. *The Telegraph»,* recogido de http://.www.telegraph.co.uk/news/uknews/1539715/Colgate-gets-the-brush-off-for-misleading-ads.html.

52 http://www.c-span.org/about/history.

53 Nielsen informa que el promedio de los estadounidenses tienen acceso a 189 canales, pero ven únicamente 17. http://www.nielsen.com/us/en/insights/news/2014/changing-channels-americans-view-just-17-channels-despite-record-number-to-choose-from.html.

54 B. Boxall (2 de diciembre de 2014), «Rancho Santa Fe ranked as state's largest residential water hog», *Los Angeles Times,* http://www.latimes.com/local/california/la-me-water-rancho-20141202-story.html.

I. Lovett (29 de noviembre de 2014), «Where grass is greener, a push to share drought's burden», *New York Times,* http://www.nytimes.com/2014/11/30/us/where-grass-is-greener-a-push-to-share-droughts-burden.html.

55 *N. del T.:* El acre equivale a poco menos de media hectárea (0,404).

56 *N. del T.:* «Capacitador de flujo» es la invención del profesor Emmet Brown en la película *Regreso al futuro* que permite realizar viajes en el tiempo.

57 http://www.flightsafety.org; K. B. Grant (30 de diciembre de 2014), «Deadly year for flying — but safer tan ever», http://www.cnbc.com/id/102301598.

58 Para una temperatura inicial de 155° Fahrenheit, la fórmula es $f(t) = 80e^{-0.08t} + 75$

59 P. Bedard (22 de junio de 2010), «Brian Lamb: C-SPAN now reaches 100 million homes», *U.S. News & World Report,* www.usnews.com/news/blogs/washington-whispers/2010/06/22/brian-lamd-c-span-now-reaches-100-million-homes, recogido el 22 de noviembre de 2010.

60 Basado en Huff, *op. cit.,* p. 48.

61 https://www.cbo.gov/sites/default/Files/113th-congress-2013-2014/workingpaper/49837-Casualties_WorkingPaper-2014_1.pdf.

62 https://www.census.gov/prod/2011pubs/12statab/defense.pdf.

63 http://www.cdc.gov/nchs/fastats/deaths.htm.

64 Basado en un ejemplo de Huff, *op. cit.,* p. 83.

65 Agradezco a mi alumna Alexandra Ghelerter este ejemplo. A. Barnett (1994), «How numbers are tricking you», tomado de http://www.sandiego.edu/statpage/barnett.htm.

66 Un término de Best.

67 A. Davidson (1 de julio de 2015), «The economy's missing metrics», *New York Times Magazine*.

68 A. Shell (2 de julio de 2015), «Wall Street weighs Fed's next move after job data», *USA Today Money,* http://americasmarkets.usatoday.com/2015/07/02/wall-street-gets-what-it-wants-in-june-jobs-count/.

69 N. D. Schwartz (3 de julio de 2015), «Jobless rate fell in June, with wages staying flat», *New York Times,* B1.

70 http://www.cdc.gov/nchs/faststats/leading-causes-of-death.htm.

CÓMO SE RECOLECTAN LAS CIFRAS

El mero hecho de que aparezca una cifra no significa que esa cifra se haya obtenido adecuadamente. Recordad, tal como indicaba al comienzo de esta parte del libro, quienes recopilan estadísticas son *personas*. Son las personas quienes deciden qué calcular y cómo calcularlo. Hay toda una panoplia de errores y sesgos que pueden incorporarse al proceso de recolección y pueden llevar a millones de personas a extraer conclusiones equivocadas. Pese a que la mayoría de nosotros jamás participaremos en el proceso de recolección de las cifras, la reflexión crítica sobre el mismo se puede aprender con facilidad y está al alcance de todos.

Las estadísticas se obtienen de diversas maneras: a partir de registros (por ej., los registros de nacimientos y fallecimientos de las instituciones del Estado, de los hospitales o de las iglesias), realizando sondeos o encuestas, mediante la observación (por ej.: contando la cantidad de vehículos eléctricos que pasan por la Plaza Mayor), o realizando inferencias (si las ventas de pañales suben, probablemente está aumentando la tasa de nacimientos). Los sesgos, las imprecisiones y los errores sin mala fe pueden suceder en cualquier momento. Una parte de lo que hay que hacer para valorar las afirmaciones que se realizan es formular preguntas como «¿de veras es posible averiguar eso?» y «¿cómo lo saben?».

Muestreo

Los astrogeólogos muestrean especímenes de rocas lunares, no revisan la Luna entera. Los investigadores no hablan con todos y cada uno de los votantes para averiguar qué candidato va en cabeza, ni anotan cada persona que entra en la sala de

emergencias para ver cuánto han tenido que esperar a ser atendidos. Hacer las cosas de ese modo no resultaría práctico o sería muy caro. Por el contrario, lo que hacen es utilizar muestras para estimar la cantidad real. Cuando las muestras se toman correctamente, las estimaciones pueden ser muy, muy precisas. En los sondeos de opinión pública, una estimación sobre qué opina el país entero sobre un tema (unos 234 millones de adultos superan la edad de veintiún años) se puede obtener entrevistando tan solo a 1.067 individuos. Las biopsias que muestrean menos de una milésima parte de un órgano pueden utilizarse para determinar en qué etapa se encuentra un cáncer.

Para que una muestra tenga utilidad, ha de ser representativa. Una muestra es representativa si cada persona u objeto del grupo que estudiamos tiene la misma probabilidad de ser escogido. Si no es así, la muestra está sesgada. Si el cáncer está únicamente en una parte de un órgano y muestreamos la zona equivocada, no podremos diagnosticarlo. Si se trata de una zona muy pequeña y tomamos quince muestras de ese punto, podríamos concluir que el órgano está completamente canceroso, cuando no es así.

No siempre podemos saber con antelación a cuánta variabilidad nos enfrentamos en las biopsias o en los sondeos de opinión. Si todos los elementos de una población fueran idénticos, solo necesitaríamos como muestra uno de ellos. Si hubiera un grupo de personas genéticamente idénticas, con idénticas personalidades y experiencias vitales, podríamos saber lo que quisiéramos sobre todos ellos observando tan solo a uno. Pero cada grupo incluye alguna heterogeneidad, alguna diferencia entre sus componentes, por tanto hemos de ser cuidadosos al muestrearlo a fin de asegurarnos de que incluimos las diferencias relevantes. (No todas las diferencias son relevantes). Por ejemplo, sabemos que los seres humanos mueren si les quitamos el oxígeno. Los humanos no diferimos en esa dimensión (aunque sí lo hagamos en cuanto al tiempo que somos capaces de sobrevivir sin oxígeno). Pero si queremos conocer cuántos kilos es capaz de levantar un humano haciendo pesas acostado en un banco, hay una enorme variación, es necesario evaluar una

sección transversal numerosa, compuesta de muchas personas diferentes, para poder alcanzar la variabilidad adecuada y un promedio estable. Necesitaríamos muestrear personas altas, personas bajas, personas gordas, personas delgadas, mujeres, niños, culturistas, aficionados al sofá, gente que consuma esteroides anabolizantes y abstemios militantes. Probablemente haya otros factores relevantes, tal como cuánto durmió la persona la noche previa a realizar la prueba, cuánto tiempo hacía que comió, si estaba enfadada o tranquila, etc. También hay aspectos que creemos que no guardan ninguna relación: si ese día el controlador aéreo del aeropuerto St. Hubert de Quebec es hombre o mujer; si un cliente al azar en un restaurante de Aberdeen fue atendido a tiempo o no. Se trata de cuestiones que pueden suponer una diferencia cuando medimos otros asuntos (el sexismo latente en la industria del transporte aéreo; la satisfacción de los clientes en los establecimientos de hostelería de noroeste), pero no tienen relación con el levantamiento de pesas acostado en un banco.

La tarea del estadístico consiste en formular un inventario de todas aquellas cuestiones que son relevantes para obtener una muestra representativa. Los investigadores deben evitar la tendencia a evaluar variables que son fáciles de identificar o sobre las que se pueden recopilar datos con facilidad, a veces los aspectos relevantes no son evidentes o son difíciles de medir. Tal como dijo Galileo Galilei: la tarea del científico es medir lo que es mensurable y hacer mensurable lo que no lo es. Es decir, algunas de las actividades más creativas de la ciencia suponen vislumbrar cómo medir algo que establece la diferencia, algo que hasta ese momento nadie había logrado medir.

Pero incluso medir e intentar controlar variables que sí conocemos supone un desafío. Supongamos que deseamos estudiar las actitudes actuales hacia el cambio climático en los Estados Unidos. Disponemos de una pequeña suma de dinero para pagar ayudantes y comprar un programa de estadística para nuestro ordenador. Vivimos en San Francisco, de modo que decidimos realizar nuestro estudio allí. Ya tenemos un primer problema: San Francisco no es representativo del resto del

estado de California y mucho menos de los Estados Unidos. Al darnos cuenta de lo anterior, decidimos realizar el sondeo en agosto, porque otros estudios muestran que es la temporada alta del turismo, personas de todo el país van a San Francisco en esa época y (eso creéis) seréis capaces de obtener una muestra transversal de los estadounidenses.

Pero, esperad un momento: ¿quienes visitan San Francisco son representativos de los demás? Estaríamos sesgados hacia aquellas personas que pueden pagarse el viaje, hacia quienes desean pasar sus vacaciones en una ciudad, por oposición a, digamos, en un parque nacional. (También estaríamos sesgados hacia las personas de izquierda, ya que San Francisco es una ciudad famosa por ser de izquierda).

En consecuencia, decidimos que no podemos costear un estudio sobre las actitudes en los Estados Unidos, sino un estudio sobre las actitudes hacia el cambio climático entre los habitantes de San Francisco. Situamos a nuestros ayudantes en Union Square para que pidan a los viandantes que respondan un breve cuestionario. Les instruimos para que busquen personas de distintas edades, razas, estilos de vestir, con y sin tatuajes, en pocas palabras, una sección transversal con variabilidad. Pero seguimos teniendo problemas: no es probable que nos encontremos con quienes están postrados en su cama, con las madres con niños pequeños, con los trabajadores de turno nocturno que duermen durante el día, ni con los cientos de miles de residentes en San Francisco que por distintas razones no pasan por Union Square, una zona de la ciudad conocida por sus tiendas caras y sus restaurantes. Enviamos a la mitad de nuestros ayudantes al Distrito de la Misión para resolver el problema de los diferentes niveles socioeconómicos que deben estar representados, pero seguimos sin resolver los demás problemas. La prueba de una muestra al azar es la siguiente: ¿Tienen todos los componentes del grupo la misma probabilidad de ser evaluados por nosotros o por nuestro equipo? En este caso la respuesta es clara: no.

Por lo tanto, buscamos una muestra aleatoria *estratificada*. Es decir, identificamos diferentes estratos o subgrupos de interés y

empleamos personas de cada uno en la misma proporción que se encuentran en el conjunto de la población. Investigamos sobre el cambio climático y descubrimos que las actitudes al respecto no parecen distribuirse en función de la raza[71], de modo que no es necesario que establezcamos submuestras basadas en la raza. Lo anterior nos beneficia, ya que puede resultar complicado u ofensivo realizar supuestos a partir de la raza y, además, ¿qué hacer con las personas mestizas? ¿En qué categoría los ponemos? ¿Creamos una categoría solo para ellos? ¿Y entonces qué pasaría? ¿Una única categoría para los estadounidenses que se identifican como blancos-negros, negros-hispanos, asiáticos-persas, etc.? Las categorías pueden acabar resultando tan específicas que supongan un problema para su análisis, ya que implican demasiados grupos diferentes. Otro obstáculo[72]: nos interesa contar con la variabilidad de la edad, pero a la gente a veces no le gusta decir su edad. Podríamos escoger personas que estén claramente por debajo de los cuarenta y otros que estén claramente por encima, pero eso dejaría fuera a quienes están al final de la treintena y al comienzo de los cuarenta.

Para solventar el problema que suponen las personas que no circulan por la ciudad durante el día, podemos decidir hacer un sondeo puerta a puerta. Pero si vamos durante la jornada laboral, perdemos a quienes están trabajando; si vamos por la noche, perdemos a los que van a las discotecas, a los trabajadores de turno nocturno, a los que van a la iglesia al servicio nocturno, a los que van a cenar a restaurantes. Una vez que hemos realizado la estratificación, ¿cómo obtenemos una muestra aleatoria de cada subgrupo? Todavía persisten los problemas que acabamos de describir, establecer los subgrupos no resuelve el problema de que dentro de cada subgrupo aún hemos de procurar identificar una representación equilibrada de todos los *demás* factores que podrían tener efecto sobre los datos. Empieza a parecer que sí que necesitaríamos tener una muestra de todas las rocas de la Luna para poder hacer un buen análisis.

Pero no desesperemos, el muestreo aleatorio estratificado es mejor que el no estratificado. Si obtenemos al azar una muestra de alumnos universitarios para describir su experiencia en la

universidad, podríamos acabar con una muestra formada solo por alumnos de las grandes universidades públicas, una muestra aleatoria los incluirá más fácilmente, ya que son muchos más. Si su experiencia universitaria es radicalmente diferente de lo que sucede en las pequeñas universidades privadas dedicadas a las humanidades, sería necesario que nuestra muestra incorpore alumnos de esas universidades, de modo que nuestra muestra estratificada asegure que también vamos a sondear a estudiantes de instituciones de distintos tamaños. Es importante diferenciar el muestreo aleatorio del muestreo por cuotas o por conveniencia: la práctica de entrevistar únicamente a personas que conocemos o a personas que nos encontramos por la calle y que tienen pinta de estar dispuestas a responder. Sin una muestra aleatoria, el sondeo muy probablemente será sesgado.



«¡Tras muestrear cada pájaro que frecuenta las aceras alrededor de este edificio, he concluido que a los pájaros lo que verdaderamente les gusta son las rosquillas!».

Por tanto, la recopilación de datos mediante muestreo es una batalla interminable para evitar las distintas fuentes de sesgo y el investigador nunca la gana totalmente. Cada vez que leemos en

algún periódico el dato de que el 71% de los británicos están a favor de algo, debiéramos reflexionar y preguntarnos: «De acuerdo, pero ¿el 71% de *qué* británicos?»[73].

Al hecho anterior hay que añadir que cualquier pregunta que hagamos sobre las personas es únicamente una muestra de todas las posibles preguntas que podríamos formularles y sus respuestas pueden ser únicamente una muestra de lo que son sus complejas actitudes y experiencias. Para empeorar las cosas, cabe que comprendan o no lo que se les pregunta y también pueden distraerse mientras responden. Además, con más frecuencia de lo que a los encuestadores les gustaría reconocer, las personas a veces dan respuestas intencionadamente erradas. Los humanos somos una especie social; muchos intentan evitar el enfrentamiento y buscan agradar, de modo que dan la respuesta que creen que el encuestador querría escuchar. Por otra parte, también existen personas marginales e inconformistas que pueden responder en falso, simplemente para asombrar al encuestador[74] o meramente para ensayar un papel de personaje rebelde y disfrutar sorprendiendo o desafiando.

Obtener una muestra sin sesgos no es fácil. Cuando llega a nuestros oídos una estadística nueva, preguntémonos: «¿Qué sesgos se habrán colado durante el muestreo?»

Las muestras nos ofrecen estimaciones sobre algo y prácticamente siempre se desvían de la cifra real, sea mucho o poco, y eso se conoce como margen de error. Podemos imaginarlo como el precio que pagamos por no tener que averiguar algo sobre todos y cada uno de los miembros de la población que estudiamos[75], o por no tener que muestrear todas las rocas de la Luna. Es indudable que también pueden producirse errores incluso cuando entrevistamos o evaluamos a todas y cada una de las personas de una población, a causa de los sesgos en los instrumentos de medida. El margen de error no se debe a los defectos subyacentes de la investigación, sino únicamente al grado de error en el procedimiento de muestreo. Por el momento vamos a ignorar esos defectos más profundos y vamos a considerar otra medida o estadístico que se encuentra

relacionado con cualquier muestra rigurosamente definida: el intervalo de confianza.

El margen de error indica cuán precisos son los resultados y el intervalo de confianza indica hasta qué punto hemos de confiar en que nuestras estimaciones se sitúen dentro del margen de error. Por ejemplo, en una encuesta estándar de dos vías (en la que se presentan dos alternativas a quienes responden), en una muestra aleatoria de 1.067 estadounidenses adultos, da lugar a un margen de error próximo al 3% en cualquiera de las direcciones (se escribe ±3%). Por lo tanto, si una encuesta muestra que el 45% de los estadounidenses apoyan al Candidato A y el 47% apoyan al Candidato B, la cifra real se sitúa en algún punto entre el 42 y el 48% para A y entre el 44 y el 50% para B. Obsérvese que las horquillas de ambos se solapan[76]. Esto significa que la diferencia de dos puntos entre el Candidato A y el Candidato B se encuentra dentro del margen de error: no podemos afirmar que uno de ellos está verdaderamente por delante del otro, se trata de una elección demasiado reñida para afirmarlo.

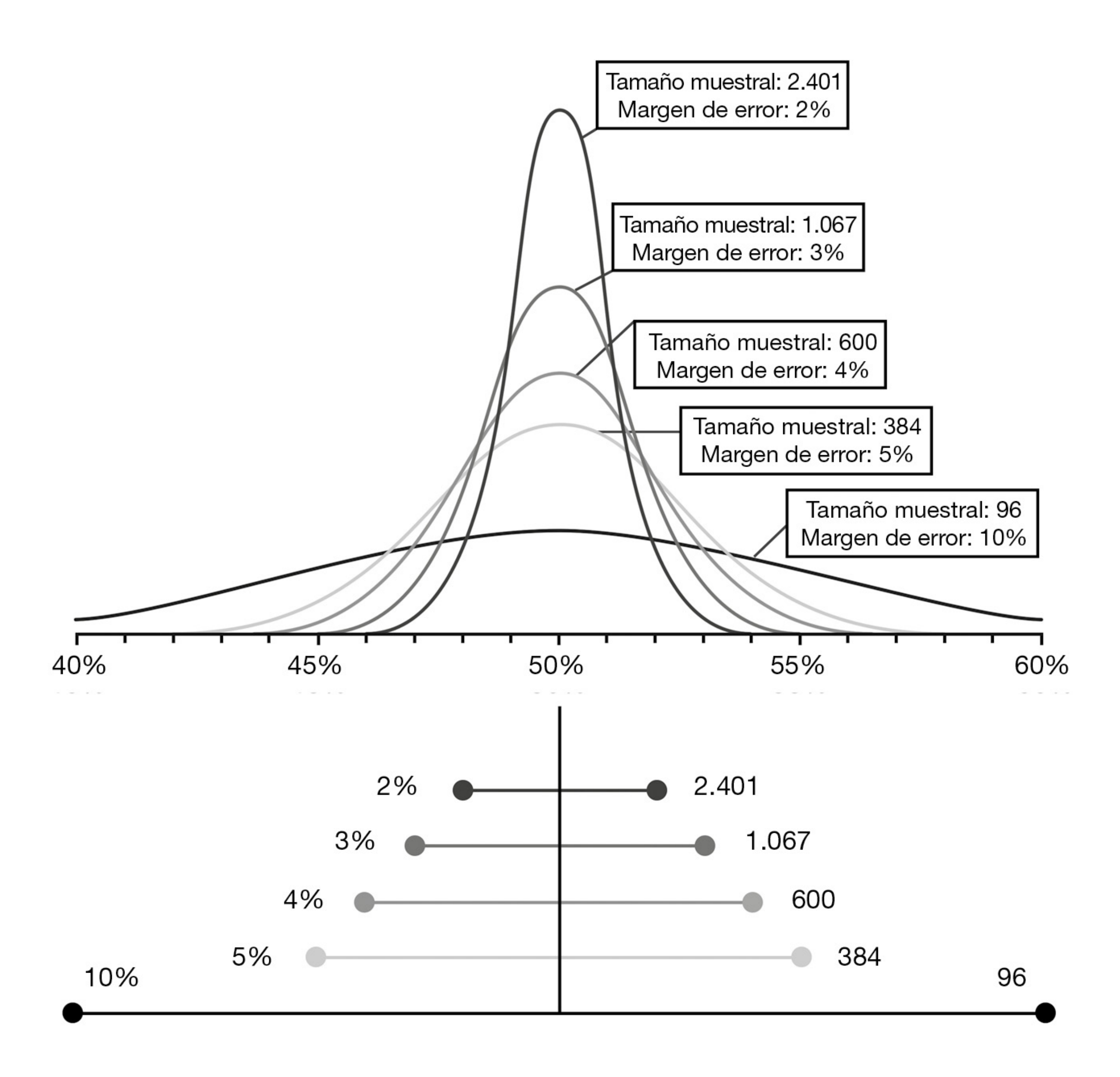
¿Hasta qué punto podemos fiarnos de que el margen de error es 3% y no más? Para responder necesitamos calcular un intervalo de confianza. En el caso concreto que he mencionado, he utilizado un intervalo de confianza del 95%. Eso significa que si realizara la encuesta cien veces, utilizando el mismo método de muestreo, noventa y cinco de esas cien veces el intervalo que obtendríamos incluiría el valor verdadero. En cinco de las cien veces[77] obtendríamos un valor situado fuera de ese rango, podría tratarse de una diferencia pequeña o de una grande; hay otras estadísticas que podrían ayudarnos a establecerlo.

Podemos situar el intervalo de confianza en cualquier valor que queramos, pero lo habitual es el 95%. Para alcanzar un intervalo de confianza menor se pueden hacer dos cosas: *incrementar* el tamaño de la muestra para un nivel de confianza dado o, para un tamaño de muestra dado, *disminuir* el nivel de confianza. Para un tamaño de muestra dado, el cambio del nivel de confianza del 95 al 99% aumentará el tamaño del intervalo. En la mayoría de los casos los gastos añadidos y los inconvenientes que se derivan

hacen que no valga la pena, dado que una multitud de circunstancias externas pueden hacer que de todos modos la opinión de las personas cambie al cabo de un día o de una semana.

Obsérvese que para poblaciones muy grandes, como la de Estados Unidos, solo necesitamos muestrear un porcentaje muy pequeño, en este caso menos del 0,0005%. Pero con poblaciones más pequeñas, como la de una empresa o una escuela, necesitaremos porcentajes mucho mayores. Para una empresa de 10.000 empleados, necesitamos una muestra de 964 (casi un 10%), para alcanzar un margen del 3% con un 95% de confianza, y para una empresa de 1.000 empleados, necesitaríamos una muestra de casi 600 (el 60%).

El margen de error y el intervalo de confianza se aplican a los muestreos de cualquier tipo, no meramente a las muestras de personas: también para muestrear la proporción de coches eléctricos en una ciudad o de células malignas en el páncreas o de mercurio en el pescado del supermercado. En la figura siguiente, se muestran el margen de error y el tamaño de la muestra para un intervalo de confianza del 95%[78].



La fórmula para calcular el margen de error[79] (y el intervalo de confianza) se encuentra en las notas al final del libro y es posible acceder a muchas calculadoras en línea que nos permiten realizarlo. Si veis que se cita una estadística y no se ofrece el margen de error, es posible calcularlo, siempre que se conozca la cantidad de personas que fueron sondeadas. En muchos casos veréis que el periodista o la organización que hace la encuesta no proporcionan esa información. Es lo mismo que si nos muestran un gráfico sin ejes, se puede mentir muy fácilmente con la estadística si no se informa del margen de error o del intervalo de confianza. Veamos un caso: mi perro Sombra es el candidato que va en cabeza para la elección de Gobernador del estado de Mississippi, un 76% de los votantes tienen intención de votarlo

frente a los demás candidatos (con un margen de error del que no se ha informado de ±76; ¡Votad por Sombra!).

Sesgos de muestreo

Cuando intentan obtener una muestra aleatoria, los investigadores a veces cometen errores al determinar si cada persona u objeto tienen la misma probabilidad de ser incluidos en la muestra.

En la elección presidencial de Estados Unidos en el año 1936 se cometió un error terrible. La revista *Literary Digest* realizó una encuesta y concluyó que el republicano Alf Landon vencería al demócrata en el poder, el presidente Roosevelt. El *Digest* había sondeado a personas que eran lectores de revistas, dueños de automóviles o que tenían teléfono, no a una muestra aleatoria. La explicación convencional, citada en muchas publicaciones académicas y populares, es que en 1936, los aspectos mencionados sesgaban decididamente a favor de los ricos, quienes tenían más tendencia a votar republicano. De hecho, según una encuesta realizada por Gallup en 1937, esta explicación convencional es errónea[80], los dueños de automóviles y de teléfonos se inclinaban más a votar por Roosevelt. El sesgo consistía en que quienes apoyaban a Roosevelt tenían menos probabilidad de participar de una encuesta. Este sesgo de muestreo fue identificado por Gallup, quien realizó su propia encuesta utilizando una muestra aleatoria y predijo correctamente los resultados. Había nacido la encuesta Gallup, que llegó a convertirse en el patrón oro de las encuestas políticas, hasta que se equivocaron al identificar el ganador de la elección presidencial de 2012 en Estados Unidos. Una investigación desveló graves problemas[81] en sus procedimientos de muestreo, que irónicamente afectaban a propietarios de teléfonos.

Al igual que las encuestas de los años treinta y cuarenta, basadas en las listas telefónicas, estaban sesgadas hacia los ricos, hoy día las encuestas basadas en los teléfonos fijos están

sesgadas hacia las personas de más edad. Todos los muestreos telefónicos se basan en el supuesto de que quienes poseen un teléfono representan al conjunto de la población; esto puede ser así o no. Muchos de quienes trabajan en Silicon Valley utilizan aplicaciones de Internet para conversar, de modo que el muestreo telefónico podría estar subrepresentando a los individuos con acceso a alta tecnología.

Si queréis mentir con la estadística y ocultar vuestros pasos, medid la estatura de las personas cerca de una cancha de baloncesto; preguntad por los ingresos cerca de una oficina de desempleo; estimad la incidencia del cáncer de pulmón limitando vuestro muestreo al entorno de una fundición. Si no desveláis cómo seleccionasteis vuestra muestra, nadie os descubrirá.

Sesgo de participación

Quienes desean participar en un estudio y quienes no lo desean pueden diferir en cuestiones importantes, como sus opiniones políticas, su personalidad y sus ingresos. De modo semejante, quienes responden afirmativamente a los anuncios para reclutar voluntarios para un estudio, vuestros voluntarios, pueden tener algún sesgo a favor o en contra de aquello que os interesa estudiar. Si pretendemos reclutar a la persona «promedio» para nuestro estudio, podríamos sesgar la participación por el mero hecho de informar con antelación sobre el objeto del estudio. Una investigación sobre las actitudes hacia el sexo estará sesgada a favor de quienes tengan mayor deseo de desvelar tales actitudes y en contra de los tímidos y los pudorosos. Un estudio sobre las actitudes políticas estará sesgado a favor de quienes están más deseosos de discutirlas. Por esta razón muchos cuestionarios, encuestas y estudios psicológicos no indican por anticipado qué es lo que se va a investigar, o enmascaran el verdadero propósito del estudio mediante una serie de preguntas irrelevantes que no se relacionan con lo que interesa al investigador.

Las personas que participan hasta el final en el estudio pueden ser diferentes que quienes lo abandonan antes de concluir. Es

posible que algunas de las personas con las que contactamos se limiten a no responder. Esto puede suponer un sesgo cuando los tipos de personas que responden a nuestra encuesta son diferentes de quienes no lo hacen, lo que da lugar a una clase de sesgo especial que se denomina sesgo de no respuesta.

Imaginemos que trabajamos para la Universidad de Harvard y deseamos demostrar que quienes se licencian allí tienden a ganar buenos sueldos ya a los dos años de graduarse. Remitimos cuestionarios a todos los que se graduaron. Ya tenemos un problema: las personas que cambiaron de dirección sin comunicarlo a Harvard, quienes estén en la cárcel o los sintecho no recibirán nuestra encuesta. Además, entre quienes respondan, los que tengan buenos sueldos y un buen recuerdo de Harvard tendrán más probabilidad de cumplimentar el cuestionario que quienes estén desempleados y resentidos. Las personas de quienes no tenemos datos contribuyen al error de no respuesta y a veces lo hacen de un modo sistemático que distorsiona los datos.

Si el objetivo de nuestro sondeo sobre el salario dos años después de terminar en Harvard es demostrar que la educación que allí se ofrece da lugar a sueldos altos, esa encuesta podría ayudarnos a demostrarlo para la mayoría de las personas. Pero un pensador crítico podría darse cuenta de que el tipo de gente que va a Harvard no es la persona promedio. Suelen ser de familias con ingresos altos y eso correlaciona con los ingresos futuros del alumno. Los estudiantes de Harvard tienden a ser ambiciosos, puede que ganaran un salario igual de alto si hubieran estudiado en una universidad menos reputada, o incluso si no hubieran ido a la universidad (Mark Zuckerberg, Matt Damon y Bill Gates son personas con éxito financiero que no terminaron sus estudios en Harvard).

Si no podemos acceder a algún sector de la población, como los militares destinados en el exterior, los sintecho o las personas recluidas en instituciones, este sesgo de muestreo se denomina *error de cobertura,* porque algunos miembros de la población que deseamos muestrear no pueden ser incluidos y, por tanto, no tienen probabilidad de ser seleccionados.

Si intentamos estimar qué proporción de gominolas[82] en un tarro son rojas, naranjas y azules, puede ocurrir que no podamos llegar al fondo del tarro. Las biopsias de los órganos a menudo se limitan al tejido que el cirujano puede recolectar y eso no es necesariamente una muestra representativa del órgano. En las investigaciones psicológicas, los sujetos experimentales frecuentemente son estudiantes universitarios que no representan a la población en general. Hay una enorme diversidad de personas en el país, con actitudes, opiniones, posiciones políticas, experiencias y estilos de vida diferentes. Aunque sería un error afirmar que todos los estudiantes universitarios son semejantes, también sería un error afirmar que representan fielmente al resto de la población.

Sesgo al informar

A veces las personas mienten cuando se les pregunta su opinión. Un licenciado por Harvard puede exagerar sus ingresos para dar la imagen de que tiene más éxito, o puede indicar lo que cree que hubiera debido ganar si no fuera por alguna circunstancia que lo impide. Sin duda, también puede ser que declare menos ingresos para que la Asociación de Antiguos Alumnos de Harvard no le pida una donación elevada. Estos sesgos pueden, o no, compensarse mutuamente. El promedio que obtenemos en una encuesta sobre los sueldos de los alumnos de Harvard es tan solo el promedio de lo que han informado, no de lo que realmente ganan. Los acaudalados puede que no tengan una idea muy clara de sus ingresos anuales porque no todo son sueldos, se incluyen muchos otros ingresos que pueden variar de año en año, como lo que se obtiene a partir de inversiones, dividendos, complementos, derechos de autor, etc.

Puede que preguntemos a la gente si han copiado en un examen o si han mentido al pagar impuestos. Puede que no crean que la encuesta es verdaderamente confidencial y no deseen informar de su conducta verdadera. (Esto supone un problema cuando se intenta estimar la cantidad de inmigrantes

ilegales en Estados Unidos que necesitan cuidados sanitarios o que son víctimas de la delincuencia; muchos tienen miedo de ir a los hospitales o a las comisarías porque temen ser denunciados a las autoridades de inmigración).

Imaginemos que queremos saber qué revistas lee la gente[83]. Podríamos preguntarles, pero tal vez quieran darnos una buena impresión, o quizá pretendan considerarse más refinados en sus gustos de lo que realmente son. Podríamos encontrar que muchas más personas declaran leer el *New Yorker* o la *Atlantic* de lo que indican sus cifras de ventas y que muchas menos personas declaran leer *US Weekly* o el *National Enquirer*[84]. Las personas no siempre declaran la verdad en las encuestas, de modo que en realidad no estaríamos midiendo lo que leen, sino su esnobismo.

Entonces organizamos un plan: vamos a ir a los hogares de la gente y vamos a ver qué revistas tienen en la sala de estar. Pero también esto implica un sesgo: no nos informa de lo que realmente leen, únicamente nos indica lo que han decidido conservar después de haberlo leído, o lo que deciden mostrar para dar una buena impresión. Saber qué revistas leen las personas es más difícil de medir que saber qué revistas *compran* (o muestran). Sin embargo, se trata de una diferencia importante, sobre todo para los anunciantes.

¿Qué factores subyacen al hecho de que una persona se identifique como mestiza? Si se ha criado en una comunidad formada por una única raza, puede que no se sienta inclinada a considerarse mestiza. Puede que *nosotros* seamos capaces de definir el mestizaje con precisión, pero eso no significa que las personas vayan a declararlo como querríamos.

Falta de estandarización

Las medidas deben ser estandarizadas. Han de ser procedimientos claros, replicables y precisos para recoger datos, de modo que cada persona que los emplea lo haga del mismo modo. Cada persona que hace la cuenta debe hacerla del mismo

modo. Consideremos el índice Gleason[85] para clasificar los tumores, está solo relativamente estandarizado, lo que significa que podemos obtener diferentes puntuaciones de Gleason y, por tanto, distintas calificaciones de la etapa del cáncer, de patólogos diferentes. (Para establecer la puntuación de Gleason se examina al microscopio una muestra de tejido y se asigna una puntuación de 2 a 10 para indicar la probabilidad de que el tumor avance.) Los psiquiatras difieren en sus opiniones sobre si cierto paciente tiene esquizofrenia o no. Los estadísticos no se ponen de acuerdo sobre qué constituye una demostración suficiente de los fenómenos psíquicos. La patología, la psiquiatría, la parapsicología y otros campos procuran generar procedimientos bien definidos para que todo el mundo pueda obtener los mismos resultados, pero casi todas las medidas presentan ambigüedades y dan pie a que surjan diferentes opiniones. Si se nos pide que nos pesemos, ¿lo hacemos vestidos o desnudos, con la cartera en el bolsillo o sin ella? Si se nos pide que midamos la temperatura de un filete en una parrilla, ¿lo hacemos en un solo punto o en varios para promediar?

Error de medida

Los participantes pueden no comprender una pregunta tal y como pretendía el investigador; puede que marquen la casilla errada en el cuestionario, o por mor de muy diversas razones podrían no dar la respuesta que pretendían. Los errores de medida suceden en todas las medidas, en todos los campos de la ciencia. Los físicos del CERN informaron de que habían medido neutrinos desplazándose a más velocidad que la luz, un descubrimiento que hubiera podido situarse entre los más importantes de los últimos cien años. Más tarde declararon que se había tratado de un error de medida[86].

Los errores de medida tienen lugar cuando cuantificamos algo. En las elecciones presidenciales de Estados Unidos del año 2000 tuvo lugar un error de medida (y de registro equivocado de las intenciones del pueblo): distintos equipos de funcionarios

contando los mismos votos obtuvieron cifras diferentes. En parte se debió a desacuerdos sobre cómo contabilizar un papel perforado o marcado, etc., problemas de definición, pero incluso cuando se pusieron en práctica instrucciones estrictas, continuaron apareciendo diferencias en el cómputo.

Todos hemos tenido la siguiente experiencia: cuando contamos céntimos de la jarra donde los guardamos, obtenemos totales diferentes si los contamos dos veces. Cuando nos subimos a la báscula del baño tres veces seguidas, obtenemos pesos diferentes. Cuando medimos el tamaño de una habitación en nuestra casa, es posible que obtengamos longitudes ligeramente diferentes en cada ocasión. Se trata de sucesos explicables: los muelles de la báscula son instrumentos mecánicos imperfectos. Cada vez que sujetamos la cinta métrica lo hacemos de un modo un poco distinto, se desliza un poco el punto desde el que empezamos a medir, leemos los milímetros incorrectamente, o la cinta no es lo suficientemente larga para medir la habitación y hemos de marcar un punto en el suelo para realizar la medida en dos o tres partes, lo que aumenta la posibilidad de un error. El propio instrumento de medida puede incorporar variabilidad (de hecho, los instrumentos de medida incluyen especificaciones sobre su precisión, cuanto más caro el instrumento, más preciso suele ser). La báscula del baño puede que sea precisa solo hasta los doscientos gramos, una báscula de correos puede serlo hasta mucho menos, unos pocos gramos.

Un censo de Estados Unidos de 1960 registró[87] sesenta y dos mujeres de edades comprendidas entre los quince y los diecinueve que tenían doce o más hijos, y una gran cantidad de viudas de catorce años. El sentido común nos indica que no puede haber muchas chicas entre los quince y los diecinueve años con doce hijos y que las viudas de catorce son muy inusuales. Alguien cometió un error. Puede que algunos de los encuestadores del censo hayan cumplimentado la casilla equivocada en el impreso de modo accidental o quizá intencionado, a fin de evitar tener que realizar largas entrevistas. O quizá un grupo de ciudadanos impacientes (o traviesos) inventó

historias extravagantes y los encuestadores del censo no se dieron cuenta.

En 2015 el equipo New England Patriots fue acusado de manipular sus balones, sacándoles aire para que resultara más fácil atraparlos. Esgrimieron el error de medida como parte de su defensa[88]; aquel día se midió la presión de los balones de ambos equipos, Patriots e Indiana Colts, tras el descanso. Los balones de los Patriots se controlaron primero y después los de los Colts. Los balones de los Colts pudieron haber estado guardados en un vestuario caluroso durante más tiempo, lo que habría provocado que se calentasen y aumentase su presión. El juzgado federal del distrito aceptó el argumento, lo que sumado a otro testimonio le llevó a decidir que no había pruebas suficientes de manipulación.

Los errores de medida también ocurren cuando el instrumento que utilizamos para medir, la báscula, la regla, el cuestionario o la prueba, no mide lo que se pretende que mida. Emplear una vara de medir para estimar el ancho de un cabello o un cuestionario sobre depresión, cuando lo que realmente nos interesa es la motivación (pueden estar relacionadas, pero no son lo mismo), puede provocar este tipo de error. Evaluar qué candidatos reciben apoyo financiero de las personas no es lo mismo que saber a quién van a votar; hay muchas personas que ofrecen contribuciones económicas a varios candidatos en la misma elección.

Se ha escrito mucho sobre distintas pruebas y encuestas que pretendían mostrar una cosa y mostraban otra. Las pruebas de CI se encuentran entre las que dan lugar a más malinterpretaciones. Se utilizan para medir la inteligencia de las personas, como si la inteligencia fuera una cualidad única; no hay tal, la inteligencia se manifiesta de distintas maneras, como la inteligencia espacial, la inteligencia artística, la inteligencia matemática, etc. Además, sabemos que los test de inteligencia están sesgados a favor de las personas blancas de clase media. Lo que habitualmente queremos saber cuando consultamos los resultados de un test de CI es hasta qué punto es apta una persona para un cierto empleo o para un programa educativo. Las pruebas de CI pueden predecir el rendimiento en esa situación, pero probablemente no

es porque las personas con CI más alto sean necesariamente más inteligentes, sino porque esas personas tienen un historial de mayores ventajas (económicas, sociales) que se muestran en el test de CI.

Si la estadística que nos encontramos se basa en una encuesta, procurad averiguar qué preguntas se utilizaron y si estas os resultan sensatas y carentes de sesgos. Con cualquier estadística es necesario intentar averiguar cómo se midió el asunto que se estudia y si las personas que recopilaron los datos eran competentes con ese tipo de medidas.

Definiciones

Cómo se define o se categoriza puede provocar enormes diferencias en la estadística que se obtiene. Este problema surge en las ciencias naturales cuando se intentan clasificar las células cancerosas o describir la pluviosidad y también en las ciencias sociales, cuando se pregunta a las personas sobre sus opiniones o experiencias.

¿Llovió hoy en el área del Gran San Luis? Depende que cómo definamos llover. Si únicamente cayó una gota de agua en los 22.911 km^2 del «Gran San Luis» (según el Departamento de Gestión y Presupuesto de Estados Unidos), ¿diremos que ha llovido? ¿Cuántas gotas han de caer sobre un área de qué extensión durante cuánto tiempo para que afirmemos que el día fue lluvioso?

El Departamento de Estadísticas Laborales de Estados Unidos utiliza dos modos diferentes para medir la inflación en función de dos definiciones distintas. Los Gastos de Consumo Personal (GCP) y el Índice de Precios al Consumo (IPC) pueden arrojar cifras diferentes. Si comparamos dos años o dos regiones del país, es necesario que nos aseguremos de que empleamos el mismo índice cada vez. Si lo que pretendemos es meramente resaltar que la inflación se elevó o disminuyó recientemente, el usuario de la estadística carente de escrúpulos utilizaría el que ofreciera datos más llamativos, en lugar de elegir el más

adecuado, fundándose en el conocimiento de las diferencias entre ambos.

¿Qué significa ser un sintecho? ¿Es alguien que duerme en la acera o que duerme en un coche? Puede ser que se trate de una persona que tiene casa, pero que no pudo llegar o que eligió no dormir allí. ¿Y qué decir de una mujer que duerme en el sofá de un amigo porque ha perdido su apartamento? ¿O de una familia que ha vendido su casa y que se aloja en un hotel durante un par de semanas, mientras terminan de arreglar la casa nueva? ¿Y de un hombre que vive feliz y contento como okupa en un almacén abandonado? Si comparamos el concepto de sintecho en distintas ciudades y estados, comprobaremos que las diversas jurisdicciones emplean definiciones diferentes. Incluso si la definición entre las distintas jurisdicciones fuera homogénea, la estadística que podemos observar puede no haber definido sintecho[89] como nosotros lo haríamos. Una de las dificultades para resolver el «problema de los sintecho» en las ciudades grandes es que no disponemos de una definición generalizadamente aceptada de qué es o quién cumple el criterio.

Cuando encontréis una noticia nueva sobre una nueva investigación, es importante estar alerta sobre cómo han sido definidos los elementos de esa investigación. Debemos juzgar si son aceptables y razonables. Este es un asunto de especial importancia cuando se trata de temas muy politizados, como el aborto, el matrimonio, la guerra, el cambio climático, el salario mínimo y las políticas de vivienda.

Nada está más politizado que la política. Una definición puede ser disputada y distorsionada para obtener ventajas en los sondeos de opinión formulando la pregunta adecuada. Imaginad que os ha contratado un candidato político[90] para que busquéis información sobre su oponente, Alicia Florrick[91]. A menos que Florrick haya conseguido entusiasmar a cada uno de los ciudadanos sobre cada uno de los temas, éstos tendrán alguna insatisfacción. Así que lo que hacemos es preguntar: «¿Hay algo con lo que esté en desacuerdo o que no le convenza, de cualquier cosa que haya dicho la candidata, incluso si piensa

votarla?». Dado que casi todo el mundo tendrá alguna insatisfacción, podréis ofrecer a vuestro jefe el dato de que «el 81% de las personas están insatisfechas con Florrick». Lo que habréis hecho es recopilar datos sobre un tema (aunque se trate de un desacuerdo menor) y apilarlos sobre otras quejas semejantes, renombrándolos como «insatisfacción». Casi suena decente.

Cosas que no pueden conocerse o que son inverificables

El término GIGO *(garbage in, garbage out)* es famoso y fue acuñado por los primeros informáticos: entra basura, sale basura. En aquellos tiempos la gente confiaba ciegamente en cualquier resultado que ofreciera un ordenador porque esos resultados parecían investidos de precisión y certeza. Si una estadística se construye a partir de un conjunto de medidas mal definidas, intuiciones, errores conceptuales, sobre simplificaciones, medidas equivocadas o estimaciones incompletas, las conclusiones resultantes serán imperfectas.

Buena parte de lo que leemos debiera despertar nuestra suspicacia. Preguntaos: ¿cabe que alguien conozca esto? Un periódico informa sobre la proporción de suicidios[92] cometidos por adolescentes gays y lesbianas. Cualquier estadística de ese tipo está condenada a carecer de sentido, dada la dificultad que supone conocer qué muertes son suicidios y qué cadáveres corresponden a homosexuales o a heterosexuales. De manera semejante debiéramos sospechar de la cifra de muertes por hambre en una región remota o de la cantidad de personas asesinadas en un genocidio en una guerra civil. Lo anterior se puso de manifiesto en las estimaciones radicalmente divergentes realizadas por los observadores sobre las bajas en el conflicto Irak-Afganistán-Estados Unidos.

El editor de una revista se vanagloria de que su revista tiene 2 millones de lectores. ¿Cómo lo sabe? No lo sabe. Supone que una proporción de las revistas vendidas es compartida por varias personas, lo denominan «tasa de difusión». Suponen que cada

revista comprada por una biblioteca es leída por una cierta cantidad de personas. Otro tanto se aplica a los libros y a los libros electrónicos. Sin duda, hay una gran variabilidad dependiendo de los títulos. Mucha gente compró *Breve historia del tiempo* de Stephen Hawking. De hecho, se dice que es el libro más vendido y menos leído de los últimos treinta años. Probablemente pocas personas se lo prestaron a otro, impresiona mucho que se vea en la sala. ¿Cuántos lectores tiene una revista o un libro? ¿Cuántos oyentes tiene un podcast? No lo sabemos. Podemos saber cuántos se vendieron o cuántas descargas hubo, eso es todo (aunque algunos avances recientes del libro electrónico probablemente van a alterar este *statu quo* establecido hace bastante tiempo).

La próxima vez que leáis que el neozelandés promedio utiliza hilo dental 4,36 veces por semana (un dato que me acabo de inventar, pero que es tan preciso como cualquier otra estimación), podéis preguntaros: ¿cómo puede alguien saber tal cosa? Si hubiera cámaras ocultas en los cuartos de baño sería otro cantar, pero lo más probable es que se trate de personas que responden cuestionarios y únicamente responden aquello que pueden recordar, o lo que prefieren pensar que es la verdad, siempre tendremos que enfrentarnos a eso.

---

71 Esto es completamente hipotético.

72 Tomado de Huff, *op. cit.,* p. 22.

73 *Ibid*.

74 Hace muchos años, el columnista de Chicago Mike Royko alentó a sus lectores a mentir a los encuestadores a pie de urna el día de las elecciones con la esperanza de que los datos inexactos y aparecer como unos tontos acabase con la práctica de los comentaristas de televisión de anunciar el resultado antes del recuento de votos. Carezco de datos sobre cuántas personas mintieron a la salida merced a la petición de la columna de Royko, pero el hecho de que las encuestas a pie de urna continúen existiendo sugiere que no fueron los suficientes.

75 Tomado de http://www.aapor.org/Education-Resources/Election-Polling-Resources/Margin-of.Sampling-Error-Credibility-Interval.aspx.

76 Es una regla de oro, pero en algunos casos este método rápido resulta impreciso. Véase N. Schenker y J. F. Gentleman (2001), «On judging the significance of differences by examining the overlap between confidence intervals», *American Statistician,* 55(3), 182-186.

77 Intencionadamente no diferencio aquí entre las estimaciones bayesiana y frecuentista de la probabilidad, una distinción que aparece en la segunda parte.

78 Imagen del margen de error tomada de la Wikipedia.

79 Para poblaciones grandes el intervalo de confianza del 95% puede estimarse como ±1,96 × sqrt [p(1-p)/n]. Para obtener un intervalo de confianza del 99%, multiplíquese por 2,58 en lugar de por 1,96. Cierto, el intervalo es *mayor* cuando tenemos más confianza (lo que debiera tener sentido; si deseamos estar más seguros de que el rango que citamos incluye el valor verdadero, necesitamos un rango mayor). Con poblaciones más pequeñas, lo primero es calcular el error típico:

sqrt [{(Proporción observada) × [I – (Proporción observada)}/tamaño de la muestra]

Entonces la amplitud del intervalo de confianza del 95% es ± el error típico.

Por ejemplo, si hemos muestreado cincuenta pasos elevados en una gran ciudad, podríamos observar que el 20% de los mismos necesitan ser reparados. Calculamos el error típico del siguiente modo:

sqrt [(0,2 × 0,8)/50] = sqrt (0,16/50) = 0,057.

En consecuencia, la amplitud de nuestro intervalo de confianza del 95% es de ± 20,057 = ±0,11 o ±11%. Por tanto el intervalo de confianza del 95% implica que el 20% de los pasos elevados de la ciudad necesitan ser reparados, con un margen de error del 11%. Para incrementar la precisión de nuestra estimación, necesitamos una muestra mayor. Si revisamos 200 pasos elevados (suponiendo que obtuviéramos el mismo dato del 20%), nuestro margen de error se reduciría a aproximadamente el 6%.

80 D. Lusinchi (2012), «"President" Landon and the 1936 *Literary Digest* poll: were automobile and telephone owners to blame?», *Social Science History,* 36(1), 23-54.

81 S. Clement (4 de junio de 2013), «Gallup explains what went wrong in 2012», *Washington Post,* https://www.washingotnpost.com/news/the-fix/wp/2013/06/04/gallup-explains-what-went-wrong-in-2012/.

http://www.gallup.com/poll/162887/gallup-2012-presidential-election-polling-review.aspx.

82 Tomado de https://ropercenter.cornell.edu/support/polling-fundamentals-total-survey-error/.

83 Elaborado a partir de un ejemplo de Huff, *op. cit.,* pág. 16.

84 *N. del T.: New Yorker* y *Atlantic* son dos revistas de perfil intelectual más alto que *Us Weekly* o *National Inquirer*.

85 Esta definición está tomada literalmente de http://www.cancer.gov/publications/dictionaries/cancer-terms?cdrid=45696, consultado el 20 de marzo de 2016.

86 F. Jordans (23 de febrero de 2012), «CERN researchers find flaw in faster-than-light measurement», *Christian Science Monitor,* http://www.csmonitor.com/Science/2012/0223/CERN-researchers-find-flaw-in-faster-than-light-measurement.

87 Tomado de R. D. De Veaux y D. J. Hand (2005), «How to lie with bad data», *Statistical Science,* 20(3), 231-238, p. 232. Citan a W. Kruskal (1981), «Statistics in society: problems unsolved and unformulated», *Journal of the American Statistical Association,* 76(375), 505-515 y A. J. Coale y F. F. Stephan (1962), «The case of Indians and the teen-age widows», *Journal of the American Statistical Association,* 57, 338-347.

88 J. Kryk (15 de mayo de 2015), «Patriots strike back with compelling explanations to refute deflate-gate chargers», *Ottowa Sun,* http://www.ottawasun.com/2015/05/14/patriots-strike-back-with-compelling-explanations-to-refute-deflate-gate-chargers.

89 Ejemplo de H. Spirer, L. Spirer y A. J. Jaffe (1998), *Misused Statistics,* 2.ª edición revisada y ampliada, Nueva York, Marcel Dekker, p. 16.

90 Ejemplo basado en el de Huff, *op. cit.,* pág. 80.

91 *N. del T.:* Alicia Florrick es la protagonista de una conocida serie de televisión sobre intrigas políticas y jurídicas, *The Good Wife*.

92 De Best (2005), *op. cit.*

PROBABILIDADES

¿Me creísteis cuando afirmé que *probablemente* pocas personas habían prestado *Breve historia del tiempo*? Utilizaba el término vagamente, como suele hacerse, pero el tema de la probabilidad matemática aborda los propios límites de lo que podemos o no conocer sobre el mundo, desde la conducta de partículas subatómicas, como los quarks y los bosones, hasta la probabilidad de que el fin del mundo tenga lugar en el curso de nuestras vidas, desde la gente que juega a la lotería hasta el intento de pronosticar el tiempo (dos esfuerzos que podrían arrojar tasas de éxito semejantes).

Las probabilidades nos permiten cuantificar sucesos futuros y constituyen una ayuda importante para la toma de decisiones racionales. Sin ellas podríamos vernos seducidos por anécdotas y narraciones. Es posible que hayáis oído a alguien decir algo parecido a: «no me voy a poner el cinturón de seguridad[93], me han contado que un tipo se mató en un accidente de automóvil *porque* lo llevaba puesto. Se quedó atrapado y no pudo salir. Si no hubiera llevado el cinturón de seguridad abrochado se habría salvado».

Vale, está bien, pero no podemos limitarnos a considerar un par de historias. ¿Cuáles son los riesgos relativos? Aunque hay unos pocos casos raros en los que el cinturón de seguridad le ha costado la vida a alguien, es mucho más probable morir cuando *no* lo llevamos. La probabilidad nos ayuda a considerar estos asuntos de modo cuantitativo.

Empleamos la palabra *probabilidad* de distintas maneras para significar cosas diferentes. Es fácil dejarnos llevar y creer que una persona quiere decir una cosa, cuando en realidad quiere decir otra y la confusión puede llevarnos a extraer conclusiones equivocadas.

Un tipo de probabilidad, la *probabilidad clásica,* se basa en la noción de simetría y de igual probabilidad[94]: un dado tiene seis caras, una moneda tiene dos caras, una rueda de ruleta tiene treinta y seis casillas (en Estados Unidos; en Europa son treinta y

siete). Si no hay defectos de fábrica o manipulaciones que favorezcan un resultado sobre otro, cada uno es igualmente probable. En consecuencia, la probabilidad de que salga un número en particular al lanzar un dado es de uno entre seis, de sacar cara al lanzar una moneda es de uno entre dos y de que salga un número en la rueda de la ruleta es de uno entre treinta y seis o treinta y siete.

La probabilidad clásica se limita a estos tipos de objetos bien definidos. En el caso clásico conocemos los parámetros del sistema y, por lo tanto, podemos calcular las probabilidades de los sucesos que produce cada sistema. Se produce un segundo tipo de probabilidad porque en la vida diaria a menudo nos interesa conocer la probabilidad de que tengan lugar otros sucesos, como la probabilidad de que un fármaco funcione con un paciente o de que los consumidores prefieran una marca de cerveza en vez de otra. En este segundo caso, necesitamos estimar los parámetros del sistema porque no los conocemos.

Para determinar ese segundo tipo de probabilidad, realizamos observaciones o hacemos experimentos y contabilizamos la cantidad de veces que obtenemos los resultados deseados. Estas se conocen como probabilidades *basadas en frecuencias*. Administramos una sustancia a un grupo de pacientes y contabilizamos cuántas personas mejoran, eso es un experimento y la probabilidad de que el fármaco funcione es simplemente la proporción de personas que mejoran (basada en la *frecuencia* del resultado deseado). Si realizamos el experimento con una gran cantidad de personas[95], los resultados estarán cerca de la probabilidad verdadera, igual que cuando se hace un sondeo de opinión pública.

Tanto la probabilidad clásica como la frecuentista[96] abordan sucesos que se repiten, que pueden reproducirse, y la proporción de tiempo durante el que podemos esperar que se produzca un resultado concreto bajo condiciones sustancialmente semejantes. (Algunos probabilistas de la línea dura sostienen que es necesario que se trate de condiciones *idénticas,* pero a mi modo de ver esto es forzar las cosas excesivamente, ya que en última instancia el universo nunca es *exactamente* el mismo debido a las variaciones al azar). Cuando realizamos un sondeo de opinión pública

entrevistando personas al azar, en realidad les preguntamos bajo condiciones idénticas, incluso cuando a algunos los interroguemos hoy y a otros mañana, siempre que no haya tenido lugar mientras tanto algún suceso importante que pudiera cambiar lo que piensan. Cuando un testigo declara en un juicio sobre la probabilidad[97] de que el ADN de un sospechoso corresponda al que se encontró en un revólver, utiliza una probabilidad *basada en frecuencias,* ya que esencialmente lo que hace es contar la cantidad de fragmentos de ADN que corresponden, frente a la cantidad de los que no lo hacen. Sacar una carta de una baraja, encontrar un aparato defectuoso en una línea de montaje, preguntar a la gente si les gusta una marca de café, son ejemplos de probabilidades clásicas o basadas en frecuencias que abordan sucesos que se reiteran, que pueden reproducirse (la carta de la baraja es clásica, el aparato y el café son frecuentistas).

Un tercer tipo de probabilidad es diferente de los dos primeros, ya que no surge de experimentos o de sucesos reproducibles, por el contrario, expresa una opinión o grado de creencia sobre cuán probable es que tenga lugar un suceso. Se denomina *probabilidad subjetiva* (un tipo de esta es la probabilidad bayesiana, llamada así por el estadístico del siglo XVIII Thomas Bayes). Cuando una amiga nos dice que hay un 50% de posibilidades de que asista a la fiesta de Michael y Julie este fin de semana está empleando probabilidad bayesiana, expresa la intensidad de la creencia de que asistirá. ¿Cuál será la tasa de desempleo el año próximo? No es posible utilizar el método frecuentista, ya que no podemos considerar el desempleo del año que viene como un conjunto de observaciones realizadas bajo condiciones idénticas o siquiera similares.

Veámoslo con un ejemplo. Cuando el meteorólogo de televisión dice que mañana hay un 30% de probabilidad de lluvia, no ha realizado una serie de experimentos en días idénticos bajo condiciones idénticas (si es que tal cosa pudiera existir), para después contabilizar los resultados. La cifra del 30% expresa el grado de creencia (en una escala del uno al cien) de que lloverá y pretende informarnos sobre si hemos de acordarnos de salir con el paraguas y las botas de agua.

Si el informe metereológico está bien calibrado, lloverá exactamente el 30% de los días que afirma que hay un 30% de

probabilidad de lluvia. Si llueve el 60% de esos días, ha subestimado mucho la probabilidad. La cuestión de la calibración es relevante únicamente con las probabilidades subjetivas.

Por cierto, volviendo a la amiga que nos dijo que había un 50% de posibilidades de que asistiera a la fiesta, un error que cometen muchos pensadores no críticos consiste en suponer que si hay dos posibilidades, ambas han de ser igualmente probables. En un experimento, los psicólogos cognitivos Amos Tversky y Daniel Kahneman describieron fiestas y otros escenarios a varias personas. Por ejemplo, en relación a una fiesta concreta se les dijo que el 70% de los invitados eran escritores y el 30% ingenieros. Si nos tropezamos con alguien que lleva un tatuaje de Shakespeare, probablemente supondríamos que se trata de uno de los escritores; si nos tropezamos con alguien que muestra las ecuaciones de Maxwell en la camiseta, correctamente supondríamos que es uno de los ingenieros. Pero si nos tropezamos con alguien que no muestra nada distintivo, ni tatuaje de Shakespeare, ni ecuaciones de Maxwell en la camiseta, ¿cuál es la probabilidad de que esa persona sea un ingeniero? En los experimentos de Tversky y Kahneman[98] las personas tendían a decir que un 50%, confundiendo aparentemente la existencia de dos posibilidades con la igualdad de sus probabilidades.

La probabilidad subjetiva es el único tipo de probabilidad disponible en situaciones prácticas en las que no hay experimentos, no existe una ecuación simétrica. Cuando un juez instruye al jurado para que emita un veredicto basándose en la existencia de «pruebas irrefutables» de la culpabilidad del acusado, cada miembro del jurado ha de decidir por sí mismo si se ha alcanzado el nivel irrefutable, sopesando las pruebas según sus propios estándares y creencias personales (los cuales probablemente sean idiosincrásicos, no objetivos).

Cuando un corredor de apuestas fija la apuesta en una carrera de caballos, está empleando probabilidad subjetiva, pese a que puede que esté informado de los datos sobre el historial de los caballos y de los jinetes y su salud, no se produce una simetría natural (es decir, no es un caso de probabilidad clásica) y no se está realizando ningún experimento (es decir, no es un caso de probabilidad basada en frecuencias). Lo mismo sucede con los

partidos de béisbol y otros eventos deportivos. Un corredor de apuestas puede afirmar que los Royals tienen un 80% de probabilidades de ganar su próximo partido, pero no está utilizando la probabilidad en un sentido matemático; es simplemente una forma de hablar que emplea, que empleamos, para darle una pátina de precisión numérica. El corredor de apuestas no puede hacer retroceder las manecillas del reloj una y otra vez para contabilizar cuántas veces ganan. Puede que haya digerido una pila de cifras o que utilice un ordenador para fundar sus estimaciones, pero, en última instancia, la cifra es una conjetura, una indicación de su nivel de confianza en su predicción. Que diferentes críticos especializados ofrezcan pronósticos distintos es un buen indicio de que es una probabilidad subjetiva[99].

Las probabilidades subjetivas nos rodean y la mayoría de nosotros ni siquiera nos damos cuenta, las encontramos en los periódicos, en las salas de reuniones, en los bares. Las probabilidades de que un Estado malévolo lance una bomba atómica en los próximos doce meses, de que las tasas de interés suban el año que viene, de que Italia gane la Copa del Mundo o de que las tropas conquisten una cierta colina, son todas ellas subjetivas, no basadas en frecuencias: se trata de sucesos únicos, no replicables. Y las reputaciones de los comentaristas especializados y de quienes se dedican a hacer pronósticos se basan en su exactitud.

Combinar probabilidades

Una de las reglas más importantes de la probabilidad es la regla de la multiplicación. Si dos sucesos son independientes, es decir, si el resultado de uno no influye sobre el resultado del otro, la probabilidad de que *ambos* sucedan se obtiene multiplicando las dos probabilidades entre sí. La probabilidad de que salga cara cuando lanzamos una moneda al aire es un medio (ya que hay dos posibilidades igualmente probables: cara y cruz). La probabilidad de sacar un naipe de corazones de una baraja es de un cuarto (ya que hay solo cuatro posibilidades igualmente probables: corazones, diamantes, tréboles y picas). Si lanzamos una moneda y sacamos

un naipe, la probabilidad de que salgan una cara y un corazón se calcula multiplicando las dos probabilidades individuales: ½ × ¼ = 1/8. Esto se denomina probabilidad conjunta.

Podemos asegurarnos de que lo anterior es cierto haciendo una lista de todos los casos posibles y contando cuántas veces obtenemos el resultado deseado:

Cara	Corazón	Cruz	Corazón
Cara	Diamante	Cruz	Diamante
Cara	Trébol	Cruz	Trébol
Cara	Pica	Cruz	Pica

Ignoro las rarísimas ocasiones en las que lanzamos la moneda y cae de canto o se la lleva una gaviota mientras está en el aire o cuando tenemos una baraja en la que solo hay tréboles.

Resolver la probabilidad conjunta de tres sucesos es semejante: sacar cara al lanzar una moneda, sacar un corazón de una baraja y que la próxima persona con la que nos crucemos tenga la misma fecha de cumpleaños que nosotros (esa probabilidad es aproximadamente 1 entre 365,24, aunque los nacimientos tienden a agruparse un poco y hay algunas fechas de nacimientos más frecuentes que otras, se trata de una aproximación razonable).

Probablemente habéis visitado alguna web en la que os formulaban una serie de preguntas de opción múltiple, como «¿en cuál de las siguientes cinco calles has vivido?» o «¿cuál de las cinco siguientes es tu tarjeta de crédito?». Esas web intentan autentificaros, para asegurarse de que sois quienes ellos creen que sois. Para hacerlo utilizan la regla de la multiplicación. Si respondéis correctamente una secuencia de seis preguntas, cada una de ellas con una probabilidad de uno entre cinco (0,2), la probabilidad de acertar por azar es tan sólo de 0,2 × 0,2 × 0,2 × 0,2 × 0,2, o 0,000064, es decir, 6 entre 100.000. No es tan riguroso como cuando se hace una prueba de ADN en un juicio, pero no está mal. (Si os estáis preguntando por qué no les basta con plantear un grupo de preguntas de respuesta corta, para que las cumplimentéis, en lugar de utilizar preguntas de opción múltiple, el

motivo es que hay demasiadas variantes de respuesta correcta, ¿dices que tu tarjeta es del Chase, del Chase Bank o de JPMorgan Chase? ¿Has vivido en la calle North Sycamore, en la calle N. Sycamore, o en la calle N. Sycamore? Ya lo vais pillando).

Cuando la probabilidad de los sucesos es influida por otros sucesos

La regla de la multiplicación sólo se aplica si los sucesos son independientes. ¿Qué sucesos no son independientes? El clima, por ejemplo. La probabilidad de que hiele esta noche *y* también hiele mañana no son sucesos independientes, el patrón meteorológico tiende a mantenerse por más de un día y aunque es posible que sucedan heladas súbitas, la mejor estimación sobre la noche de mañana es la temperatura de la noche de hoy. *Podríamos* calcular la cantidad de noches al año en las que la temperatura baja del punto de congelación, pongamos que son 36 en la zona donde vivís, y después anunciar que la probabilidad de que hiele esta noche es 36 sobre 365, más o menos el 10%, pero de ese modo no tendríamos en cuenta la relación de dependencia. Si afirmamos que la probabilidad de que hiele dos noches seguidas en invierno es de 10% × 10% = 1% (aplicando la regla de la multiplicación), estaríamos subestimando la probabilidad, ya que los sucesos de cada noche no son independientes; el pronóstico del tiempo para mañana es influido por el de hoy.

La probabilidad de un suceso también puede sufrir la influencia de la muestra concreta que consideremos. La probabilidad de que hiele esta noche resulta evidentemente afectada por la región del mundo que consideremos. Esa probabilidad es mayor en el paralelo cuarenta y dos que en el paralelo diez. La probabilidad de encontrar alguien que mida más de dos metros de estatura es mayor en un equipo de baloncesto que si vamos a un bar frecuentado por jinetes de carreras. El subgrupo de personas u objetos que consideremos influye sobre las estimaciones de probabilidad.

Probabilidades condicionales

Cuando valoramos afirmaciones basadas en estadísticas a veces nos desorientamos al examinar un grupo entero de personas al azar, cuando lo que realmente deberíamos considerar es un subgrupo. ¿Cuál es la probabilidad de enfermar de neumonía? No muy alta. Pero si supiéramos más sobre una cierta persona o un caso en particular, la probabilidad podría elevarse o disminuir. Esto se conoce como *probabilidad condicional*.

Planteamos dos preguntas diferentes:

1. ¿Cuál es la probabilidad de que una persona seleccionada al azar tenga neumonía?
2. ¿Cuál es la probabilidad de que una persona no seleccionada al azar, sino que muestre tres síntomas (fiebre, dolor muscular y congestión pectoral), tenga neumonía?

La segunda pregunta implica probabilidad condicional. Se denomina así porque no estamos considerando todas las condiciones posibles, sino únicamente las personas que cumplen la condición especificada. Incluso si no hacemos cuentas, podemos adivinar que la probabilidad de neumonía es mayor en el segundo caso. Es indudable que podríamos plantear la pregunta de modo que la probabilidad de padecer neumonía fuera inferior que para una persona seleccionada al azar:

1. ¿Cuál es la probabilidad de que una persona no seleccionada al azar, sino que ha dado negativo en tres pruebas de neumonía seguidas, que disfruta de un sistema inmune particularmente robusto y que hace unos minutos ha ganado la Maratón de Nueva York tenga neumonía?

De modo semejante, la probabilidad de que padezcamos cáncer de pulmón no es independiente de nuestro historial familiar. La probabilidad de que el camarero traiga kétchup a la mesa no es independiente de lo que hayamos pedido. Es posible calcular la probabilidad de que una persona seleccionada al azar desarrolle cáncer de pulmón en los próximos diez años, o la probabilidad de que el camarero traiga kétchup a una mesa concreta entre todas las mesas. Pero felizmente sabemos que esos sucesos son dependientes de otras conductas, lo que nos permite reducir la población que estudiamos a fin de obtener estimaciones más

precisas. Por ejemplo, si tanto nuestro padre como nuestra madre tuvieron cáncer de pulmón, necesitaríamos calcular la probabilidad de padecer cáncer de pulmón valorando a otras personas de ese grupo seleccionado, gente cuyos padres y madres hayan tenido cáncer de pulmón. Si nuestros padres no tuvieron cáncer de pulmón, debemos considerar como subgrupo relevante el de las personas que carecen de historial familiar en ese sentido (y probablemente obtendremos unos resultados diferentes). Si deseamos conocer la probabilidad de que el camarero nos traiga kétchup, deberíamos considerar únicamente las mesas en las que los clientes hayan pedido hamburguesas o patatas fritas, no aquellas en las que han pedido tartar de atún o pastel de manzana.

Ignorar la relación de dependencia de los sucesos (suponer su independencia) puede traer consecuencias graves en el ámbito legal. Uno de esos casos fue el de Sally Clark[100], una mujer de Essex, en el Reino Unido, quien fue llevada a juicio por el homicidio de su segundo hijo. Su primer hijo había muerto en la primera infancia y su muerte había sido atribuida al SMS (Síndrome de Muerte Súbita o muerte prematura). Los fiscales argumentaron que la probabilidad de tener dos bebés que mueran de SMS es tan remota que debía haber asesinado al segundo niño. El testigo de la acusación, un pediatra, citó un estudio que afirmaba que el SMS sucede en 1 de 8.543 muertes infantiles. (La especialización del Dr. Meadow en pediatría no le convierte en un estadístico o un epidemiólogo especializado, este tipo de confusión es la base de muchas decisiones equivocadas y la comentamos en la tercera parte de este libro; un especialista en un campo no es automáticamente un especialista en otro ámbito aparentemente relacionado).

Si lo consideramos con más profundidad, podríamos poner en cuestión la cifra de las 8.543 muertes. ¿Cómo lo saben? El SMS es un diagnóstico por exclusión, es decir, no existe una prueba clínica que pueda realizarse para concluir que una muerte se produjo por SMS. Por el contrario, si los médicos son incapaces de encontrar la causa y han descartado todo lo demás, lo etiquetan como SMS. No ser capaz de encontrar algo no significa que no haya sucedido, de modo que resulta verosímil que algunas de las muertes atribuidas al

SMS en realidad fueran provocadas por causas menos misteriosas, como envenenamiento, asfixia, problemas cardíacos, etc.

Sin embargo, por seguir el argumento, aceptemos que el SMS es la causa de 1 entre 8.543 muertes de bebés, tal como declaró el testigo experto. Además añadió que la probabilidad de que dos muertes por SMS sucedieran en la misma familia era de 1/8.543 × 1/8.543 o 1 entre 73 millones. («¿Coincidencia? *No* lo creo», debe haber gritado el fiscal en su alegato). Este cálculo, esta aplicación de la regla de la multiplicación, supone que las muertes son independientes, pero podrían no serlo. Lo que quiera que causara la muerte repentina del primer hijo de la señora Clark podría estar presente para ambos niños al estar ambos en el mismo hogar: dos factores ambientales relacionados con el SMS son ser fumador pasivo y poner a dormir al bebé sobre su barriga. O tal vez el primer niño sufría un defecto congénito de algún tipo, este tendría una probabilidad relativamente alta de aparecer en el genoma del segundo niño (los hermanos comparten el 50% de su ADN). Siguiendo esta línea de pensamiento, había una probabilidad del 50% de que el segundo niño muriese a causa de un factor como ese, y entonces la señora Clark no parece tan asesina de niños. Finalmente su marido encontró pruebas en los archivos del hospital de que la muerte del segundo niño tuvo una causa microbiológica. La señora Clark fue liberada, pero sólo tras pasar tres años en prisión por un crimen que no había cometido.

Hay un tipo de notación especial para las probabilidades condicionales. La probabilidad de que el camarero nos traiga kétchup, dado que pedimos una hamburguesa, se escribe así:

P(kétchup | hamburguesa)

donde la barra vertical | significa *dado*. Obsérvese que esta notación elimina muchas palabras de la descripción lingüística, de modo que la expresión matemática es sucinta.

La probabilidad de que el camarero nos traiga kétchup, dado que acabamos de pedir una hamburguesa y hemos pedido kétchup, es la siguiente:

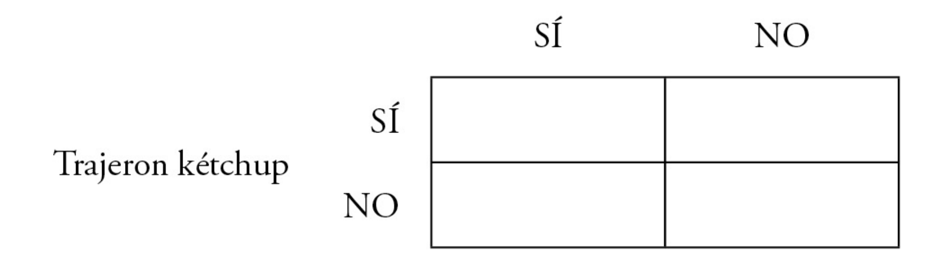
P(kétchup | hamburguesa ∧ pidió)

donde ∧ se lee como *y*.

Imaginar las probabilidades condicionales

La incidencia relativa de la neumonía[101] en los Estados Unidos está en torno al 2% al año; seis millones de personas, sobre los 324 millones del país, son diagnosticadas anualmente (seguro que también hay muchos casos sin diagnosticar, además de individuos que padecen más de una neumonía al año, pero vamos a ignorar esos detalles por ahora). Por lo tanto, la probabilidad de que cualquier persona seleccionada al azar tenga neumonía es del 2%. Pero podríamos ser mucho más certeros si sabemos algo sobre esa persona en concreto. Si vamos al médico tosiendo, con congestión y fiebre, ya no somos una persona seleccionada al azar, somos alguien en la consulta del médico que muestra esos síntomas. Se puede actualizar metódicamente nuestra creencia de que algo es verdad (de que tenemos neumonía) a la luz de pruebas nuevas. Para hacerlo aplicamos el teorema de Bayes para estimar la probabilidad condicional[102]: ¿cuál es la probabilidad de que tenga neumonía *dado* que presento el síntoma x? Este tipo de actualización se vuelve más sofisticada cuanta más información tengamos. ¿Cuál es la probabilidad de que tenga neumonía *dado* que presento estos síntomas y *dado* que tengo un historial familiar en ese sentido y *dado* que acabo de pasar tres días con alguien que estaba enfermo? Las probabilidades aumentan sin parar.

Podemos calcular las probabilidades utilizando la fórmula del teorema de Bayes (se encuentra en el Apéndice), pero es fácil visualizar y calcular probabilidades condicionales con la tabla de doble entrada, en la que se describen todos los escenarios posibles: pedimos o no pedimos una hamburguesa y nos trajeron o no nos trajeron el kétchup:



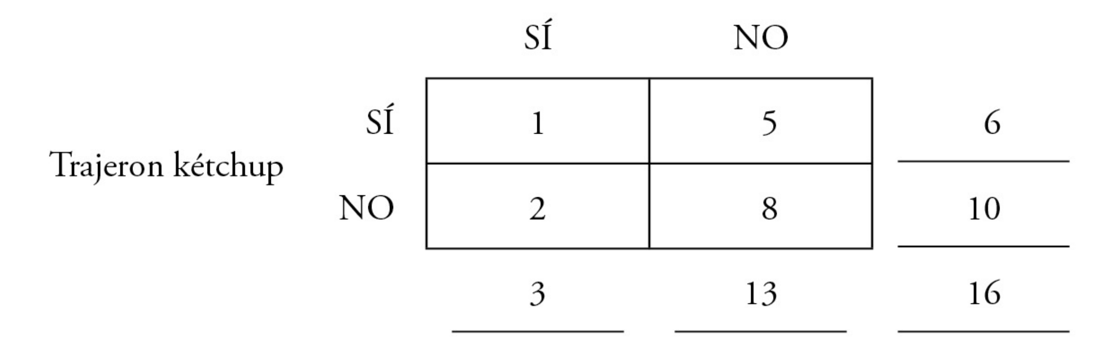
		Pedimos hamburguesa	
		SÍ	NO
Trajeron kétchup	SÍ		
	NO		

Seguidamente, basándonos en experimentos o en la observación, cumplimentamos los valores, es decir, las frecuencias de cada suceso. De los dieciséis clientes que observamos en el restaurante, hubo un caso de alguien que pidió una hamburguesa y le trajeron kétchup y dos en los que no lo trajeron. Estos son los datos de la columna de la izquierda de la tabla:

		Pedimos hamburguesa	
		SÍ	NO
Trajeron kétchup	SÍ	1	5
	NO	2	8

También observamos que cinco personas no pidieron hamburguesa y les trajeron kétchup y ocho a quienes no se lo trajeron. Esas son las entradas de la columna del lado derecho.

Seguidamente sumamos las filas y las columnas:



		Pedimos hamburguesa		
		SÍ	NO	
Trajeron kétchup	SÍ	1	5	6
	NO	2	8	10
		3	13	16

Ahora es fácil calcular las probabilidades. Si deseamos saber cuál es la probabilidad de que nos traigan kétchup *dado* que pedimos una hamburguesa, comenzamos con lo dado. Esa es la columna vertical de la izquierda.

		Pedimos hamburguesa		
		SÍ	NO	
Trajeron kétchup	SÍ	1	5	6
	NO	2	8	10
		3	13	16

Tres personas pidieron hamburguesas, ese es el total al final de la columna. Entonces, ¿cuál es la probabilidad de que traigan kétchup *dado* que pedimos hamburguesa? Miramos la casilla «SÍ trajeron kétchup» de la columna «SÍ pedimos hamburguesa y la cifra es 1. La probabilidad condicional P(kétchup | hamburguesa) es de uno sobre tres. Es fácil imaginar la lógica del asunto: tres personas pidieron hamburguesa; a una le trajeron kétchup y a dos no. Para hacer este cálculo ignoramos la columna de la derecha.

Podemos emplear este cálculo para estimar cualquier probabilidad condicional, incluyendo la probabilidad de que traigan kétchup cuando *no* pedimos una hamburguesa: trece personas no pidieron hamburguesa, a cinco les trajeron kétchup, así que la probabilidad es de cinco sobre trece, o aproximadamente un 38%. En ese restaurante concreto es más probable que te traigan kétchup si no pides una hamburguesa que si la pides. (Ahora desencadenad vuestro pensamiento crítico: ¿por qué podría ocurrir algo así? Quizá los datos corresponden a personas que pidieron patatas fritas. Tal vez todas las hamburguesas que sirven ya llevan kétchup).

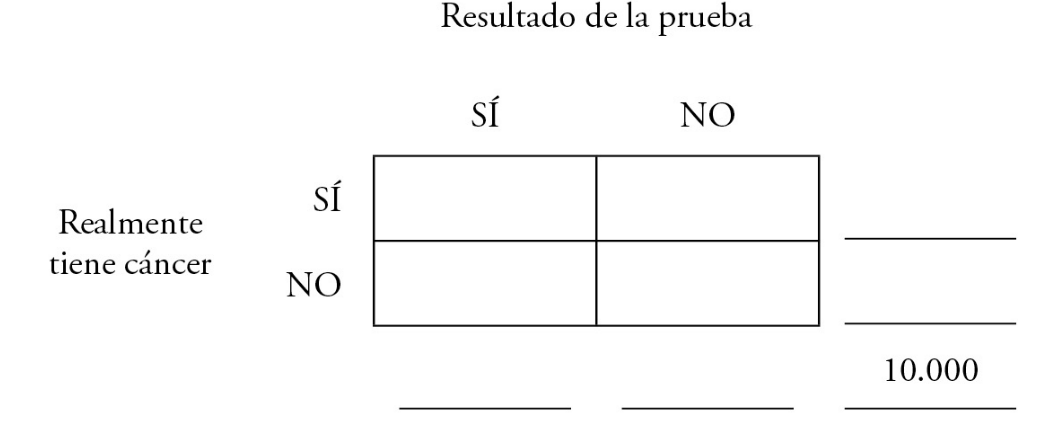
Toma de decisiones en medicina

Este modo de concebir las probabilidades condicionales resulta útil para la toma de decisiones en medicina. Cuando realizamos una prueba clínica que nos indica si padecemos una enfermedad, ¿cuál es la probabilidad de que realmente estemos enfermos? No es el 100 por cien, ya que las pruebas no son perfectas, dan lugar a falsos positivos (cuando indican que padecemos la enfermedad pero no es así) y a falsos negativos (cuando indican falsamente que no estamos enfermos).

La probabilidad de que una mujer padezca cáncer de mama[103] es del 0,8%. Si realmente lo tiene, la probabilidad de que una mamografía lo detecte es solo del 90%, ya que la prueba no es perfecta y no detecta algunos de los casos. Si la mujer no tiene cáncer de mama, la probabilidad de que obtenga un resultado positivo es del 7%. Pues bien, supongamos que una mujer seleccionada al azar obtiene un resultado positivo, ¿cuál es la probabilidad de que realmente tenga cáncer de mama?

Empezamos por diseñar la tabla de doble entrada y rellenar las posibilidades: que la mujer realmente tenga cáncer o que no lo tenga y que la prueba lo detecte o no. Para facilitar los cálculos[104] (para asegurarnos de que manejamos números enteros), vamos a suponer que contabilizamos 10.000 mujeres.

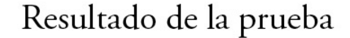
Ese es el total de la población, así que los colocamos en la esquina inferior izquierda de la figura, fuera de las casillas.

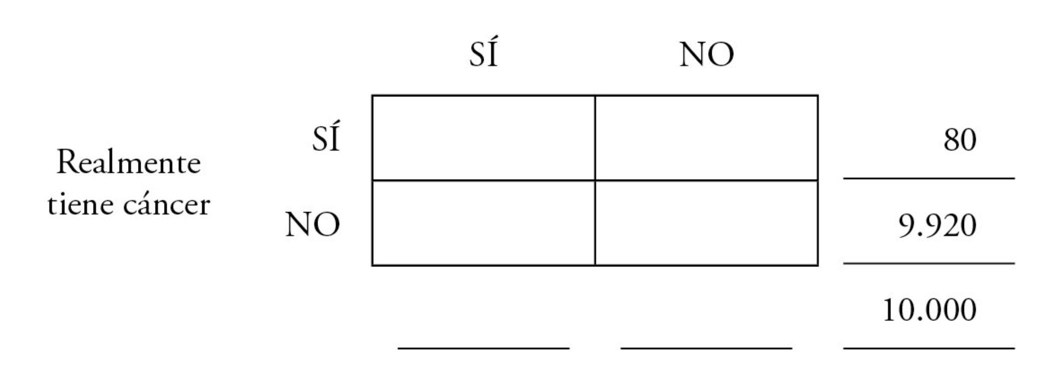


Al contrario de lo que hicimos con el ejemplo de la hamburguesa y el kétchup, rellenamos primero los márgenes, ésa es la información que nos dan. La probabilidad de tener cáncer de mama es de 0,8 o 80 entre 10.000 personas. Esa es la cifra que

colocamos en el margen de la fila superior. (Todavía no sabemos cómo rellenar las casillas, pero enseguida lo sabremos). Y ya que la fila ha de sumar 10.000, sabemos que el margen de la fila inferior ha de ser igual a

10.000 – 80 = 9.920





Realmente tiene cáncer		SÍ	NO	
	SÍ			80
	NO			9.920
				10.000

Se nos ha indicado que la probabilidad de que la prueba arroje un resultado positivo *si* realmente existe un cáncer es del 90%. Ya que la probabilidad alcanza hasta el 100 por cien, la probabilidad de que la prueba *no* arroje un resultado positivo si realmente existe un cáncer ha de ser 100 menos 90%, o sea, 10%. De las 80 mujeres que realmente tienen cáncer de mama (el margen de la fila superior), sabemos que el 90% de las mismas obtendrán un resultado positivo en la prueba (90% de 80 = 72) y un 10% obtendrá un resultado negativo (10% de 80 = 8). Lo anterior es todo lo que necesitamos saber para poder rellenar las casillas de la fila superior.

Resultado de la prueba

Realmente tiene cáncer		SÍ	NO	
	SÍ	72	8	80
	NO			9.920
				10.000

Todavía no estamos listos para calcular la respuesta a preguntas como «¿cuál es la probabilidad de que tenga cáncer de mama, teniendo en cuenta que he tenido un resultado positivo en la prueba?», ya que necesitamos saber cuántas personas obtendrán un resultado positivo en la prueba. La pieza que falta para completar el rompecabezas es la descripción original: 7% de ellas = 694,4. (Lo redondeamos a 694). Esto significa que 9.920 – 694 = 9.226, que es lo que va en la casilla inferior derecha.

		Resultado de la prueba		
		SÍ	NO	
Realmente tiene cáncer	SÍ	72	8	80
	NO	694	9.226	9.920
		766	9.234	10.000

Por último, sumamos las columnas.

Si os encontráis entre los millones de personas que creen que obtener un resultado positivo en una prueba significa que definitivamente estáis enfermos, os equivocáis. La probabilidad condicional de tener cáncer de mama dado un resultado positivo en la prueba es la casilla superior izquierda dividida por el total del margen de la columna izquierda, o 72/766. La buena noticia es que, *incluso con una mamografía positiva, la probabilidad de tener realmente un cáncer de mama es* del 9,4%. Esto es así porque se trata de una enfermedad relativamente rara (menos del 1 por 1.000) y la prueba para detectarla es imperfecta.

Resultado de la prueba

		SÍ	NO	
Realmente tiene cáncer	SÍ	72	8	80
	NO	694	9.226	9.920
		766	9.234	10.000

Las probabilidades condicionales no funcionan hacia atrás

Estamos habituados a ciertas simetrías de las matemáticas desde la enseñanza secundaria: si x = y entonces y = x; 5 + 7 = 7 + 5. Pero algunos conceptos no funcionan así, como vimos en el comentario previo sobre los valores de la probabilidad (si la probabilidad de una falsa alarma es del 10%, eso no significa que la probabilidad de un diagnóstico correcto sea del 90%).

Consideremos la siguiente estadística:

> Se venden diez veces más manzanas en los supermercados que en los puestos callejeros.

Una breve reflexión nos permitirá entender que esto no significa que sea más probable que el día que queremos comprar una manzana la encontremos en el supermercado: puede que el supermercado tenga diez veces más clientes que el puesto callejero, pero incluso así, puede que no logre abastecer la demanda. Si vemos una persona al azar por la calle con una manzana y no disponemos de información sobre dónde la compró, la probabilidad de que la haya comprado en un supermercado es mayor de que la comprase en un puesto callejero. Podemos preguntarnos en términos de probabilidad condicional: ¿cuál es la probabilidad de que esa persona la comprara en un supermercado, dado que lleva una manzana?

P (fue en un supermercado | encontró una manzana que comprar)

Eso no es lo mismo que nos gustaría saber si nos apeteciera comer una manzana Honeycrisp:

P (encontró una manzana que comprar | fue en un supermercado)

Esta misma asimetría surge de modos muy diversos en todo tipo de estadísticas. Si leemos que suceden más accidentes de automóvil a las siete de la tarde[105] que a las siete de la mañana, ¿qué significa? El lenguaje del propio enunciado es ambiguo; puede significar que estamos considerando la probabilidad de que fueran las siete de la tarde dado que ocurrió un accidente, o la probabilidad de que ocurriera un accidente dado que eran las siete de la tarde. En el primero de los casos estamos considerando todos los accidentes y determinando cuántos tuvieron lugar a las siete de la tarde. En el segundo caso consideramos cuántos automóviles circulan a las siete de la tarde y observamos la proporción de los mismos que sufre accidentes, ¿o no?

Tal vez a las siete de la tarde están circulando más automóviles que en cualquier otro momento del día y haya menos accidentes por cada mil automóviles. Eso supondría más accidentes a las siete de la tarde que a cualquier otra hora, simplemente porque hay una mayor cantidad de vehículos circulando. La *tasa* de accidentes nos ayuda a determinar cuál es la hora más segura para conducir.

De modo semejante también habréis oído decir que la mayoría de los accidentes tienen lugar a tres millas del hogar. Tal cosa no sucede porque esa zona sea más peligrosa *per se,* sino porque la mayoría de los desplazamientos que realiza la gente son cortos, de modo que una distancia de tres millas del hogar es la que se recorre con más frecuencia. En la mayoría de los casos estas dos interpretaciones del enunciado no serán equivalentes:

P(7 p.m. | accidente) ≠ P(accidente | 7 p.m.)

Las consecuencias de tal confusión no son meramente teóricas: muchos juicios en los tribunales han dependido de una aplicación errónea de la probabilidad condicional, confundiendo la dirección de lo que se conoce. Un experto forense puede calcular correctamente que la probabilidad de que la sangre que aparece en el escenario de un crimen coincida por azar con el tipo sanguíneo del acusado

es de sólo un 1%. Esto *no* es lo mismo que afirmar que hay sólo un 1% de probabilidades de que el acusado sea inocente. El experto forense nos está indicando la probabilidad de una coincidencia en el tipo de sangre, *suponiendo* que el acusado sea inocente:

P(coincide la sangre | inocente)

O, expresado sencillamente, «la probabilidad de que se diera una coincidencia de la sangre, si el acusado fuera inocente». Eso no es lo mismo que la cifra que realmente queremos conocer: ¿cuál es la probabilidad de que el acusado sea inocente, dado que sí hay una coincidencia en la sangre?

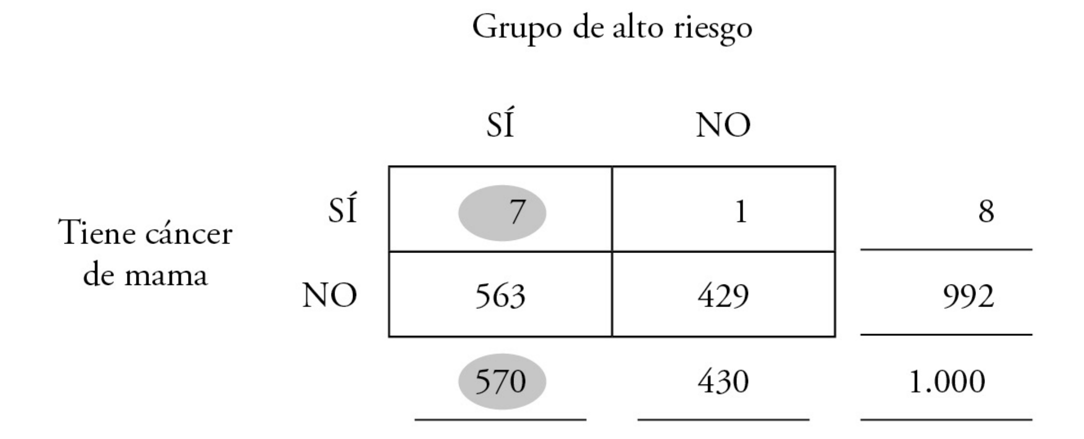
P(coincide la sangre | inocente) ≠ P(inocente | coincide la sangre)

Muchos ciudadanos inocentes han acabado en la cárcel por esta confusión. Y muchos pacientes han tomado decisiones equivocadas sobre el tratamiento médico a realizar porque pensaban erróneamente que:

P(resultado prueba positivo|cáncer) = P(cáncer|resultado prueba positivo)

Y no son sólo los pacientes, los médicos cometen este error constantemente (en una investigación el 90% de los médicos consideraron dos probabilidades diferentes como si fueran iguales)[106]. Los resultados pueden ser terribles. Un cirujano convenció a noventa mujeres[107] de que se amputasen pechos sanos si se encontraban en un grupo de alto riesgo. Se había dado cuenta de que el 93% de los cánceres de mama aparecen en mujeres que pertenecen a ese grupo de alto riesgo. Dado que una mujer tiene cáncer de mama, hay una probabilidad del 93% de que pertenezca a ese grupo: P(grupo de alto riesgo|cáncer de mama) = 0,93. Utilizando una tabla de doble entrada para una muestra de 1.000 mujeres típicas y añadiendo la información de que el 57% de las mujeres se sitúan en ese grupo de alto riesgo y que la probabilidad de que una mujer padezca cáncer de mama es de 0,8% (como mencionamos anteriormente), podemos calcular P(cáncer de mama|grupo de alto riesgo), que es la estadística que

una mujer debe conocer antes de aceptar someterse a una cirugía (redondeo las cifras sobre el número entero más próximo):



Grupo de alto riesgo

		SÍ	NO	
Tiene cáncer de mama	SÍ	7	1	8
	NO	563	429	992
		570	430	1.000

La probabilidad de que una mujer tenga cáncer, dado que se encuentre en el grupo de alto riesgo, no es del 93%, como erróneamente creyó el cirujano, sino sólo de 7/570 o 1%. El cirujano sobrestimó el riesgo de cáncer casi cien veces por encima del riesgo real, y las consecuencias fueron terribles.

Las tablas de doble entrada pueden parecer un ejercicio un poco extraño, pero cuando las elaboráis lo que estáis haciendo es pensamiento crítico y científico, disponiendo las cifras para visualizarlas mejor y facilitar los cálculos. El resultado de esos cálculos nos permite cuantificar las distintas partes de un problema, nos ayuda a adoptar decisiones más racionales, fundadas en pruebas. Las tablas de doble entrada son tan potentes que resulta sorprendente que no enseñen a elaborarlas a todo el mundo en la escuela.

Pensar sobre las estadísticas y los gráficos

La mayoría de nosotros tenemos dificultades para representarnos mentalmente las probabilidades y las estadísticas y para detectar los sutiles patrones que surgen en las tablas complejas de cifras. Preferimos las imágenes vívidas, las ilustraciones y las narraciones. Cuando tomamos decisiones, tendemos a sobreestimar esas ilustraciones y narraciones, comparadas con la información estadística; también tendemos a representar e interpretar erróneamente los gráficos.

Muchos de nosotros nos sentimos intimidados por las cifras y aceptamos sin preguntar los datos que nos ofrecen. Esto puede llevarnos a decisiones equivocadas y a conclusiones incorrectas. También tendemos a aplicar el pensamiento crítico únicamente a aquello con lo que no estamos de acuerdo. En la actual era de la información, los pseudo-hechos pasan por ser reales, la información errónea puede ser indiferenciable de la información verdadera y las cifras están en el centro de cualquier afirmación o decisión importante. La mala estadística está por doquier. Como afirma el sociólogo Joel Best[108], no es simplemente porque los demás sean unos mentirosos redomados. Las malas estadísticas son elaboradas por personas, a menudo sinceras y bienintencionadas, que no están pensando críticamente sobre lo que dicen.

El mismo temor que impide a la gente analizar las estadísticas les impide prestar la atención adecuada a las cifras de los gráficos, a las denominaciones de los ejes y a lo que todos ellos nos cuentan. El mundo está plagado de coincidencias y sucesos extraños que pueden ocurrir, pero el mero hecho de que dos sucesos cambien al mismo tiempo, no significa que uno haya causado al otro o siquiera que estén relacionados mediante un *tercer factor x*. Las personas que aceptan esas asociaciones o coincidencias muestran una mala comprensión de la probabilidad, de las causas y los efectos y del papel del azar en el desarrollo de los acontecimientos. Sin duda, es posible montar una historia sobre cómo el descenso en la cantidad de piratas durante los últimos trescientos años y la coincidente elevación de la temperatura global indican seguramente que los piratas eran esenciales para mantener bajo control el calentamiento global. Pero sería solo un razonamiento chapucero, una interpretación errónea de las pruebas. A veces quienes fomentan este tipo de lógica descabellada son conscientes de ello y esperan que no nos demos cuenta, a veces ellos mismos se lo creen. Pero ahora estáis preparados para no picar.

---

93 Este ejemplo procede de Best (2012), y de mi amigo de la infancia Kevin.

94 El principio de simetría puede representarse genéricamente como la inclusión de casos en los que los resultados no son igualmente probables pero mantienen un orden, como sucede cuando una moneda está trucada para que salga cara dos tercios de las veces, o cuando una rueda de ruleta tiene unas casillas mayores que otras.

95 También podríamos realizar el experimento muchas veces con una cantidad pequeña de personas, en cuyo caso debiéramos esperar obtener cifras diferentes. En este caso, la verdadera probabilidad de que el fármaco funcione se situará en un punto próximo al promedio (la media) de las cifras obtenidas en todos los experimentos, pero uno de los axiomas de la estadística es que cuanto mayores son las muestras, más exactos son los resultados.

96 La probabilidad clásica puede concebirse de dos modos: el empírico y el teórico. Si lanzamos una moneda al aire o sacamos cartas de un mazo que ha sido barajado, cada vez que lo hacemos es como un ensayo en un experimento que podría continuar indefinidamente. En teoría podríamos hacer que miles de personas lanzaran monedas al aire o sacaran cartas de una baraja durante varios años y acumular los resultados para obtener la proporción de ocasiones en que suceden los distintos resultados, como que «salga cara» o «salga cara tres veces seguidas». Esta sería una probabilidad *empíricamente derivada*. Si creemos que la moneda no está trucada (o sea, que no hay ningún defecto de fabricación que provoque que caiga más a menudo de un lado que del otro), no necesitamos hacer el experimento, ya que a largo plazo debiera salir cara la mitad de las veces (probabilidad = 0,5) y alcanzamos esa conclusión *teóricamente,* basándonos en el supuesto de que hay dos resultados igualmente probables. Podríamos hacer un experimento semejante con los naipes y determinar empírica y teóricamente que la probabilidad de sacar una carta de corazones es de una entre cuatro (probabilidad = 0,25) y que la probabilidad de sacar el cuatro de tréboles es de una entre cincuenta y dos (probabilidad ≅ 0,2).

97 C. G. G. Aitken y F. Taroni (2004), *Statisticas and the Evaluation of Evidence for Forensic Scientists,* 2.ª ed., Chichester, RU, John Wiley & Sons.

98 A. Tversky y D. Kahneman (1974), «Judgement under uncertainty: heuristics and biases», *Science,* 185(4157), 1124-1131.

99 Para una discusión y un tratamiento más formal, véase G. R. Iversen (1984), *Bayesian Statistical Inference,* Thousand Oaks, CA, Sage y las referencias que allí se citan.

100 Agradezco a mi alumna Alexandra Ghelerter este ejemplo. Véase también R. Nobles y D. Schiff (2007), «Misleading statistics within criminal trials», *Medicine, Science and the Law,* 47(1), 7-10.

101 http://www.nytimes.com/health/guides/disease/pneumonia/prognosis.html.

102 El teorema de Bayes es:

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}$$

103 Este párrafo y estos comentarios son una cita literal de W. Krämer y G. Gigerenzer (2005), «How to confuse with statistics or the use and misuse of conditional probabilities», *Statistical Science,* 20(3), 223-230.

104 ¿Cómo sabemos qué cifra elegir? A veces es necesario hacerlo por ensayo y error. Pero también podemos descubrir la solución. Ya que la probabilidad es 0,8%, u ocho personas entre mil, si decidimos desarrollar la tabla para 1.000 mujeres, acabaremos con ocho en uno de los cuadrantes y es lo cierto, pero más adelante vamos a tener que multiplicarlo por el 90%, lo que nos va a dar un decimal. No hay nada de malo, es meramente menos conveniente trabajar con decimales para la mayoría de las personas. Aumentar nuestra población en un orden de magnitud de 100 nos proporciona números enteros, pero estaríamos utilizando cifras mayores de las que necesitamos. En realidad no tiene importancia, ya que lo que buscamos son probabilidades y en cualquier caso vamos a tener que dividir una cifra por otra para obtener el resultado.

105 ¿Todavía confuso? Si a las siete de la tarde hubiera ocho veces más automóviles en circulación que a las siete de la mañana, la cantidad bruta de accidentes sería mayor a las siete de la mañana, pero eso no significa necesariamente que la *proporción* de accidentes en relación a la cantidad de automóviles sea mayor. *Esta* es la estadística relevante: no cuántos accidentes suceden a las siete de la mañana, sino cuántos accidentes por cada mil automóviles en circulación. Esta última formulación cuantifica el riesgo. Este ejemplo se ha modificado a partir del que se cita en Huff, *op. cit.,* p. 78 y se comenta en Krämer y Gigerenzer (2005).

106 Citado en Spirer, Spirer y Jaffe *op. cit.,* p. 197: W. C. Thompson y E. L. Schuman (1987), «Interpretation of statistical evidence in criminal trials», *Law and Human Behavior,* 11(167).

107 De Spirer, Spirer y Jaffe, *op. cit.,* señalado por primera vez en R. Hastie y R. M. Dawes (1988), *Rational Choice in an Uncertain World,* Nueva York, Harcourt Brace Jovanovich.

El informe original sobre el cirujano apareció en G. McGee (6 de febrero de 1979), «Breast surgery before cancer», *Ann Arbor News,* p. B1 (reimpreso del *Bay City News*).

108 Best, *op. cit.,* p. 184.

PARTE II

EVALUAR LAS PALABRAS

Una mentira que es una verdad a medias es siempre la más negra de las mentiras.

Lord Tennyson

¿CÓMO LO SABEMOS?

Somos una especie narrativa y social que se deja influir fácilmente por las opiniones de los demás. Tenemos tres modos de adquirir información: podemos descubrirla por nosotros mismos, podemos asimilarla implícitamente o nos la pueden contar explícitamente. Buena parte de lo que sabemos sobre el mundo forma parte de esta última categoría, alguien, en algún momento, nos ha contado un hecho o hemos leído al respecto, de modo que lo conocemos solo por referencias. Nos basamos en lo que los expertos nos cuentan.

Nunca he visto un átomo de oxígeno o una molécula de agua, pero hay una gran cantidad de bibliografía que describe experimentos meticulosos y que me lleva a creer que existen. De manera semejante, no he verificado personalmente que los estadounidenses aterrizaron en la Luna, que la velocidad de la luz es de 300.000 kilómetros por segundo, que la pasteurización mata las bacterias o que los humanos normalmente tienen veintitrés cromosomas. No he comprobado personalmente que el ascensor de mi edificio ha sido diseñado correctamente y recibe el mantenimiento correcto o que mi médico realmente fue a una Facultad de Medicina. Nos basamos en expertos, certificaciones, licencias, enciclopedias y libros de texto.

Pero también hemos de basarnos en nosotros mismos, en nuestra inteligencia y nuestra capacidad de razonar. Los mentirosos redomados que pretenden robar nuestro dinero o que votemos contra nuestros propios intereses intentan aturdirnos con pseudohechos, confundirnos con cifras que carecen de fundamento, o distraernos con información que considerada con atención no es pertinente. Se disfrazan de expertos.

El antídoto es analizar los enunciados que nos vamos encontrando igual que analizamos los gráficos y las estadísticas. Las habilidades necesarias para hacerlo no superan las

capacidades de la mayoría de los chavales de catorce años. Se enseñan en las Facultades de Derecho y de Periodismo, a veces en las Escuelas de Negocios y en los Grados de Ciencias, pero no a los demás, que somos quienes más las necesitamos.

Si os gusta asistir a películas de tema jurídico o leer artículos de periodismo de investigación, muchas de esas habilidades os resultarán familiares, recuerdan a las evaluaciones que se hacen en los tribunales. Los jueces y los jurados valoran declaraciones opuestas e intentan descubrir lo que hay de cierto en ellas. Existen reglas codificadas sobre qué constituye una prueba válida; en los Estados Unidos, los documentos que no han sido autenticados en general no se aceptan, ni tampoco los testimonios basados en rumores, aunque hay excepciones.

Supongamos que alguien nos recomienda una página web que afirma que escuchar música de Mozart durante veinte minutos al día nos hace más inteligentes. Otra web nos dice que eso no es cierto. Buena parte del problema es que el cerebro humano a menudo toma decisiones basándose en consideraciones emocionales, para después procurar justificarlas. El cerebro es una potente máquina de fabricar autojustificaciones. Sería estupendo creer que lo único que hemos de hacer es escuchar una música maravillosa durante veinte minutos para situarnos a la cabeza de la curva del CI. Evaluar enunciados como ese requiere un esfuerzo, probablemente más prolongado que el que nos lleva escuchar *Eine Kleine Nachtmusik,* pero resulta necesario para evitar extraer conclusiones equivocadas. Incluso los más inteligentes pueden ser engañados. Steve Jobs retrasó el tratamiento de su cáncer de páncreas[109] porque siguió el consejo (ofrecido en ciertos libros y páginas web) de que un cambio en la dieta podría curarlo. Cuando se dio cuenta de que la dieta no estaba funcionando, el cáncer había avanzado demasiado para poder ser tratado.

No siempre es posible determinar la veracidad o la exactitud de las fuentes. Consideremos el epigrama que abre la primera parte:

> Lo que desconoces no es lo que va a meterte en problemas, es lo que das por sentado pero no lo está.

Vi esta frase al inicio de la película *The Big Short*[110], donde se atribuía a Mark Twain y tuve la sensación de haber visto antes la cita; Al Gore también la utilizó nueve años antes en su película *Una verdad incómoda*[111], atribuyéndola al mismo autor. Pero al corroborar los datos para escribir *La mentira como arma* no pude encontrar pruebas de que Mark Twain dijera tal cosa. La atribución y la propia cita son ejemplos excelentes de aquello contra lo que nos previene la cita. Los directores, los escritores y los productores de ambas películas no hicieron sus deberes, lo que daban por sentado resultó no estarlo en absoluto.

Una breve búsqueda en la red dio lugar a un artículo en *Forbes* que afirma que es una cita apócrifa[112]. El autor, Nigel Rees, cita *Respectfully Quoted* (Citando con respeto), un diccionario de citas[113] recopilado por la Biblioteca del Congreso de los Estados Unidos. Ese volumen apunta varias formulaciones[114] de la frase en *Everybody's Friend, or Josh Billing's Encyclopedia and Proverbial Philosophy of Wit and Humor* (El amigo de todos o la enciclopedia de Josh Billings y su filosofía proverbial del ingenio y del humor) (1874). «Ahora queda claro», escribe Rees, «Mark Twain es un humorista más conocido que 'Josh Billings', de modo que la cita acaba siendo atribuida al primero».

Rees prosigue:

> Y no únicamente a él. En el debate presidencial de 1984, Walter Mondale aportó lo siguiente: «me acuerdo de una frase que Will Rogers dijo sobre Hoover[115]. Dijo: «lo que me preocupa no es lo que desconoce, lo que me inquieta es lo que da por sentado pero no lo está"».

¿Quién está en lo cierto? Con asuntos complicados como este, a menudo es conveniente consultar a un experto. Pregunté a Gretchen Lieb, investigadora bibliotecaria en Vassar, quien trabaja como enlace con el Departamento de Lengua Inglesa y me proporcionó el siguiente análisis esclarecedor:

Las citas son un asunto traicionero. Realmente son el equivalente literario de las estadísticas en cuanto a mentiras, mentiras descaradas, etc. Además, las citas más antiguas son casi como traducciones de otro idioma, en tanto que constituyen interpretaciones más que versiones literales, en particular en el caso de este círculo, ya que sus autores escribían en una especie de dialecto fantástico, al estilo de Huckleberry Finn, que es difícil de leer y en algunos casos resulta directamente inquietante.

Podría revisar muchas otras compilaciones de citas, como la de Oxford, etc., pero eso significaría ambientarse en exceso en el siglo XX.

¿Conoces la obra del HathiTrust? Es el corpus de libros que surge de las bibliotecas de investigación que soportan Google Books, y es una mina de oro, especialmente para textos impresos antes de 1928.

He aquí la atribución que se hace a Josh Billings en *Respectfully Quoted* (disponemos de este volumen en formato de libro electrónico; ¡no he necesitado moverme de mi silla!) y cita el *Oxford Dictionary of Quotations,* que tiendo a utilizar más que el Bartlett:

«"El problema de las personas no es que haya cosas que no conocen, es que conocen demasiadas cosas que no son ciertas". Atribuida a Josh Billings (Henry Wheeler Shaw) por el *Oxford Dictionary of Quotations,* 3.ª ed., p. 491 (1979). No es posible corroborarlo en sus escritos, aunque en *Everybody's Friend, or Josh Billing's Encyclopedia and Proverbial Philosophy of Wit and Humor (El amigo de todos o la enciclopedia de Josh Billings y su filosofía proverbial del ingenio y del humor)* (1874) pueden encontrarse ideas semejantes. Con la ortografía original corregida: "Qué poco sé de lo que esté seguro" (p. 502), "La sabiduría no consiste en conocer más novedades, sino en conocer menos falsedades" (p. 430), "Honestamente, creo que es mejor no saber nada que saber cosas que no son ciertas" (p. 286)».

Por cierto, en cuanto a la atribución de Walter Mondale a Will Rogers, *Respectfully Quoted* indica que no se encuentra en la obra de Rogers.

Anexo un enlace al libro de Billings, donde puedes buscar la frase «ain't so» (no lo está) y hacerte una idea de lo que hay: http://hdl.handle.net/2027/njp.32101067175438.

No se puede corroborar. Además, si buscas en Mark Twain, encontrarás que este escritor de compendios/enciclopedias cita un colega humorista y sabelotodo, Mark Twain, como su correspondiente de más confianza, de

> modo que conversan e intercambian entre ellos aforismos ingeniosos o, como hubiera dicho Billings, «afuerismos»[116]. A saber quién dijo qué.
>
> Me harta mucho cuando la gente, en especial los políticos, citan a Mark Twain o a Will Rogers. Me digo qué poco te conocemos, H. L. Mencken[117]. Hay pocas mentes críticas como la suya en estos tiempos. Pobre Josh Billings, ser el segundo humorista más famoso de su época le pone en una situación precaria cien años después.

De modo que tenemos el sorprendente caso de una cita que parece haber sido descaradamente inventada, tanto en su contenido como en su atribución. La idea básica se encontraba en Billings, pese a que no está claro si la idea viene de él, de Mark Twain o de su colega Bret Harte[118]. Will Rogers entra en la jugada porque, bueno, es el tipo de cosa que *suena* como algo que él hubiera podido decir.

La cita que encabeza la segunda parte me fue ofrecida por un conocido que la recordaba erróneamente como:

> La más negra de las mentiras es una verdad parcial que te lleva a una conclusión falsa.

Resultaba verosímil. Hubiera sido muy propio de Tennyson dar colorido a un nombre abstracto y mezclar la metafísica con lo práctico. Encontré la verdadera cita («Una mentira que es una verdad a medias es siempre la más negra de las mentiras») cuando me detuve a corroborar los enunciados de este libro. Así es la vida, como diría Kurt Vonnegut[119].

Cuando nos encontramos enunciados contradictorios, podemos elegir justificadamente y basándonos en las pruebas sobre lo que es cierto. Examinamos los enunciados nosotros mismos y tomamos una decisión, actuando como nuestro propio juez y nuestro jurado. Como parte del proceso habitualmente suele tener sentido buscar la opinión de los expertos. ¿Cómo podemos identificarlos?

---

109 J. Swaine (21 de octubre de 2011), «Steve Jobs “se arrepintió de haber intentado durante tanto tiempo vencer al cáncer mediante la medicina

alternativa”», http://www.telegraph.co.uk/technology/apple/8841347/Steve-Jobs-regretted-trying-to-beat-cancer-with-alternative-medicine-for-so-long.html.

110 *N. del T.: The Big Short* es una película del año 2015 que fue titulada en español *La gran apuesta,* trata de la crisis financiera de 2007-2010.

111 *N. del T.: An Inconvenient Truth* es un documental del año 2006 sobre el calentamiento global y la campaña electoral de Al Gore, titulado *Una verdad inconveniente* en español.

112 N. Rees (13 de agosto de 2009), «Policing Word abuse», *Forbes,* http://www.forbes.com/2009/08/12/nigel-rees-misquotes-opinions-rees.html.

113 S. Platt (ed.) (1989), *Respectufully Quoted,* Washington, D.C., Library of Congress. A la venta gracias al apoyo de Docs., USGPO.

114 J. Billings (1874), *Everybody's Friend, os Josh Billing's Encyclopedia and Proverbial Philosophy of Wit and Humor,* Hartford, CT, American Publishing Company.

115 *N. del T.:* Will Rogers fue un humorista, vaquero, actor, comentarista y estadounidense de la primera mitad del siglo XX. Herbert Hoover fue presidente de Estados Unidos entre 1929 y 1933, en su mandato se produjo la gran crisis económica de 1929.

116 *N. del T.:* En el original se emplea una ortografía antigua de «aforismo» en inglés, «afferism», para sugerir el estilo arcaico de escritura de Billings.

117 *N. del T.:* Henry Louis Mencken, periodista, escritor y crítico social estadounidense de la primera mitad del siglo XX, fue conocido por su lucha a favor de los derechos civiles, su erudición y su escepticismo.

118 *N. del T.:* Bret Harte, escritor y poeta estadounidense de la segunda mitad del siglo XIX, famoso por sus crónicas sobre la vida de los pioneros.

119 *N. del T.:* Se trata de una alusión a la conocida novela de Vonnegut *Matadero 5,* en la que el narrador suele emplear la frase «and so it goes», que aquí traducimos como «así es la vida».

IDENTIFICAR AL EXPERTO

Lo primero que hay que hacer al evaluar un enunciado emitido por alguna autoridad es preguntarse quién o qué les otorga esa autoridad. Si la autoridad se deriva de haber sido testigo de algún suceso, ¿cuán creíbles son como testigos?

No cabe duda de que autoridades venerables pueden estar equivocadas. El gobierno de Estados Unidos se equivocaba en relación a la existencia de armas de destrucción masiva (ADM) en Irak en los primeros años de este siglo y, en un contexto menos cargado políticamente, los científicos creyeron durante muchos años que los humanos tenemos veinticuatro pares de cromosomas, en lugar de veintitrés[120]. Considerar qué opinan las autoridades reconocidas no es el último paso para evaluar un enunciado, pero es un buen paso inicial.

Los expertos se pronuncian de dos maneras y resulta esencial poder diferenciarlas. El primer modo se produce cuando revisan hechos y pruebas, los sintetizan y establecen una conclusión basada en las pruebas. En ese proceso comparten con nosotros las pruebas, por qué son relevantes y cómo contribuyen a alcanzar la conclusión. Así es como se supone que funciona la ciencia, como trabajan los tribunales y el mejor modo de adoptar decisiones en los negocios, el diagnóstico médico y las estrategias militares.

El segundo modo de manifestarse de los expertos es cuando comparten sus opiniones. Son humanos, como el resto de nosotros, y sienten inclinación por las narraciones, por atar los cabos sueltos de sus introspecciones, por preguntarse: ¿qué pasaría si...? y por las ideas aún no comprobadas. No hay nada de malo en ello, algunas buenas ideas que pueden ponerse a prueba tienen su origen en este tipo de pensamiento asociativo, pero esto no debe confundirse con un argumento lógico, basado en pruebas. Los libros y los artículos de divulgación que escriben

los especialistas y los científicos a menudo incluyen ese tipo de especulaciones desbocadas, y los compramos porque nos impresiona la especialización y el talento retórico del autor. Pero cuando las cosas se hacen correctamente, el escritor debe alzar el velo de su autoridad y permitirnos mirar tras él, revelándonos al menos algunas de las pruebas.

El término *experto* habitualmente se reserva a personas que se han sometido a una instrucción especial, que han dedicado mucho tiempo a desarrollar su pericia (por ej., doctores en medicina, pilotos, músicos o atletas) y cuyas competencias o conocimientos se consideran altos en comparación con las de los demás. Como tal, la pericia es una valoración social, comparamos las habilidades de una persona con el nivel de habilidad de otros. La pericia es relativa. Einstein era un experto en física hace sesenta años, hoy probablemente no sería considerado así, si aún estuviera vivo y no hubiera incorporado a sus conocimientos lo que saben Stephen Hawking y muchos otros físicos. La pericia también supone un continuo. Aunque John Young es tan solo una de las doce personas que han caminado sobre la Luna, no sería riguroso afirmar que el capitán Young es un *experto* en caminar sobre la Luna, pese a que sepa más al respecto que casi cualquier otra persona de este mundo.

Las personas con una instrucción semejante y con niveles de pericia parecidos no necesariamente están de acuerdo e, incluso cuando lo están, no siempre aciertan. Miles de analistas financieros expertos realizan predicciones completamente equivocadas sobre los valores de las acciones en bolsa y un pequeño número de novatos puede acabar acertando. Es conocido que todas las empresas discográficas británicas rechazaron la demo de los Beatles y un joven productor sin experiencia en música pop, George Martin, les firmó el contrato con EMI. Xerox PARC, los inventores del ordenador con interfaz gráfica, no creían que los ordenadores personales tuvieran ningún futuro; Steve Jobs, quien carecía por completo de experiencia en los negocios, pensó que se equivocaban. El éxito de los recién llegados en esos campos tiende a explicarse porque el valor de las acciones y los gustos en música pop son muy

impredecibles y caóticos. Esas cosas pasan, así que el asunto no es que los expertos nunca se equivoquen, lo que ocurre es que en términos estadísticos es más probable que acierten.

Los expertos han dicho «eso nunca funcionará» a numerosos inventores e innovadores, un ejemplo clásico son los hermanos Wright y otros aspirantes a inventar el vuelo a motor. Los hermanos Wright no llegaron a terminar la enseñanza media y carecían de instrucción formal en aeronáutica o en física. Muchos especialistas con titulaciones oficiales afirmaban que nunca sería posible conseguir que volase un objeto más pesado que el aire. Los Wright eran autodidactas y su perseverancia les llevó a convertirse en verdaderos expertos, lograron construir una aeronave más pesada que el aire y demostraron que los demás expertos estaban equivocados. La historia de Michael Lewis, *Moneyball,* muestra cómo es posible que alguien gane a los expertos desafiando la sabiduría convencional y aplicando la lógica y el análisis estadístico a un problema antiguo; el entrenador del equipo A de Oakland, Billie Beane, creó un equipo competitivo empleando estadísticas de rendimiento de cada jugador, datos que otros equipos no valoraban, y consiguió clasificar a su equipo para las eliminatorias finales dos años consecutivos, lo que elevó sustancialmente el valor del equipo.

Los expertos a menudo ostentan certificaciones, o tienen titulación superior, o son reconocidos por otras autoridades. Un mecánico certificado por el fabricante de Toyota puede considerarse un experto en Toyota. El mecánico autodidacta de la esquina puede que tenga la misma competencia y puede ser mejor y más barato. La cuestión es simplemente que las probabilidades no son tan favorables y puede resultar difícil que lo averigüemos por nosotros mismos. Es una mera cuestión de promedios: el mecánico promedio certificado por Toyota sabe más sobre arreglar nuestro Toyota que el mecánico promedio independiente. Sin duda hay excepciones y resulta necesario que pongamos en juego nuestra lógica para tratar estos asuntos. Conozco un mecánico de Mercedes que trabajó para un concesionario de Mercedes durante veinticinco años y estaba entre sus mecánicos mejor valorados y calificados. Quería hacer

más corto su trayecto hasta el trabajo y ser su propio jefe, así que decidió abrir su propio taller. Sus treinta y cinco años de experiencia (en la época que lo conocí) le conferían más experiencia que a la mayoría de los mecánicos más jóvenes del concesionario. Otro caso: un trabajador independiente puede especializarse en reparaciones que un concesionario raramente realiza, como el ajuste y recalibrado de la transmisión. Es preferible que un independiente que realiza esa reparación cinco veces al mes reconstruya nuestro diferencial, en lugar de que lo haga alguien que trabaja en un concesionario y que probablemente únicamente lo hizo una vez en la escuela de formación profesional. Es lo mismo que suele decirse sobre los cirujanos: si necesitas uno, es mejor que sea un médico que ha realizado la misma operación que tiene que hacerte doscientas veces y no una o dos, y no importa lo bien que resultaran ese par de operaciones.

En la ciencia, la tecnología y la medicina el trabajo de los expertos se publica en revistas revisadas por expertos (en seguida abordaré este tema) o mediante las patentes. Pueden haber recibido premios como el Nobel, la Orden del Imperio Británico o la Medalla Nacional de la Ciencia. En los negocios los expertos pueden haber tenido experiencias como dirigir o fundar una empresa, o acumular una fortuna (Warren Buffett, Bill Gates). También hay distinciones menores, el vendedor del mes, el mecánico del año, los premios de la comunidad al «mejor» (por ej., el mejor restaurante mexicano, el mejor instalador de tejados).

En las artes y las humanidades los expertos puede que ostenten cargos en la universidad, o su pericia puede ser reconocida por quienes tienen cargos en las universidades o en el gobierno, o por un panel de expertos. Estos paneles de expertos habitualmente se forman cuando se solicitan recomendaciones de ganadores previos de un premio o por observadores bien situados, así es como se constituyen los paneles de selección de «genios» candidatos a los premios Nobel y MacArthur.

Cuando los practicantes de las artes y las humanidades ganan un premio[121], como el Nobel, el Pulitzer, el Kennedy Center Honors (premio de Honor del Centro Kennedy), el premio de

Música Polaris, el Juno, los National Book Awards (Premios Nacionales del Libro), concluimos que son expertos en su campo. Los premios otorgados por colegas expertos son particularmente útiles para establecer la pericia en un campo. La ASCAP (Asociación Americana de Compositores, Autores y Editores), una asociación cuyos miembros son exclusivamente letristas de canciones, compositores y editores de música, presenta premios que han sido votados por sus miembros; el premio es relevante, ya que quienes lo otorgan constituyen un panel de expertos. Los premios Grammy y de la Academia también son decididos mediante votación por expertos de la industria musical y cinematográfica, respectivamente.

Puede que estéis pensando, «un momento, en esos premios siempre hay cuestiones relacionadas con la política y con las inclinaciones personales. Mi actor/cantante/escritor/bailarín favorito nunca ha ganado ese premio, y me apuesto lo que sea que es tan bueno como quien lo ganó este año». Pero esa es una cuestión diferente. El sistema de premios en general está sesgado para que cada ganador merezca el premio, lo que no es lo mismo que decir que cada persona que lo merece lo ha ganado. (Recordad los comentarios previos sobre las asimetrías). Habitualmente quienes reciben el reconocimiento de buena fe de un premio respetable son expertos. (Una vez más hay excepciones, como cuando se otorgó el Grammy de 1990, que posteriormente fue retirado, a cantantes en *playback* como Milli Vanilli; o cuando se otorgó el Premio Pulitzer a la reportera del *Washington Post,* Janet Cooke, y se le retiró dos días después cuando se descubrió que su historia ganadora era un fraude. El novelista Gabriel García Márquez dio la puntilla al afirmar que Cooke hubiera debido ganar el Premio Nobel de *Literatura*). Cuando un experto es culpable de fraude, ¿niega esto su pericia? Tal vez. Sin duda tiene un impacto sobre su credibilidad, ahora que sabemos que han mentido una vez, debiéramos precavernos de que no lo repitan.

La pericia es típicamente restringida

El Dr. Roy Meadow, el pediatra que testificó en el caso de la acusada de infanticidio Sally Clark, no era un experto en estadística médica ni en epidemiología. *Formaba parte* de la profesión médica y el fiscal que lo subió al estrado sin duda esperaba que el jurado supusiera que era un experto. William Shockley recibió el Premio Nobel de Física por ser uno de los tres inventores del transistor. Algo más tarde promovió opiniones racistas radicales que tuvieron eco, probablemente porque la gente supuso que si era lo bastante inteligente para ganar un Nobel, debía saber cosas que otros desconocen. Gordon Shaw, que «descubrió» el hoy desacreditado efecto Mozart[122], fue un médico sin formación en ciencias de la conducta; la gente probablemente se imaginó, igual que con Shockley: «es médico, debe ser muy listo». Pero la inteligencia y la pericia tienden a ser específicas de un ámbito, al contrario de la creencia generalizada según la que la inteligencia es una capacidad singular, unificada. El mejor mecánico de Toyota del mundo puede ser incapaz de diagnosticar qué pasa con nuestro VW y el mejor especialista en derecho tributario puede no ser capaz de ofrecer el mejor consejo en un juicio por ruptura de contrato. Lo más probable es que un médico no sea la persona más indicada a quien consultar sobre ciencias sociales.

Reservamos un lugar especial en nuestro corazón (pero esperemos que no en nuestra mente racional) para los actores que utilizan la imagen del personaje que representan para anunciar ciertos productos. Por más que Sam Waterston resultase plenamente creíble como el fiable y ético fiscal del distrito Jack MacCoy en *Ley y Orden,* como actor carece de especiales conocimientos sobre finanzas e inversiones, aunque sus anuncios para TD Ameritrade eran convincentes. Una generación antes, Robert Young, adorado como *Marcus Welby* en televisión, hizo anuncios para Sanka. Actores como Chris Robinson *(General Hospital)* y Peter Bergman *(All My Children)* anunciaron Vicks 44; a causa de la normativa de la FTC (*Federal Trade Commission,* Comisión Federal de Comercio), los actores debían realizar un descargo de responsabilidad que acabó siendo muy conocido como eslogan: «No soy médico, pero interpreto a

uno en televisión». Según parece, los telespectadores crédulos confundían la autoridad del personaje de las series de televisión con su autoridad en el mundo real de la medicina.

La jerarquía de las fuentes

Algunas publicaciones consultan a expertos reales más frecuentemente que otras y existe una jerarquía de fuentes de información. Algunas fuentes son sistemáticamente más fiables que otras. En el ámbito académico los artículos que han sido revisados por expertos son generalmente más fiables que los libros y los libros editados por las principales editoriales son en general más fiables que los libros autoeditados (ya que las principales editoriales tienen mayores probabilidades de revisar y editar los textos y tienen un mayor incentivo financiero para hacerlo). Los periódicos premiados, como el *New York Times,* el *Washington Post* y el *Wall Street Journal* han ganado su reputación siendo sistemáticamente más rigurosos en las noticias que ofrecen. Se esfuerzan por obtener verificaciones independientes de cualquier noticia. Si un cargo del gobierno les dice algo, buscan a otro que lo corrobore. Si un científico afirma algo, toman contacto con otros científicos que no tengan relación con el descubrimiento para conocer sus opiniones independientes. A veces se equivocan; incluso los reporteros del *Times* han llegado a inventar noticias y el «defensor del lector» publica diariamente las erratas. Algunas personas, entre las que se encuentra Noam Chomsky[123], han afirmado que el *Times* es un órgano de propaganda, que ofrece noticias sobre el gobierno de los Estados Unidos sin la necesaria cuota de escepticismo. Pero una vez más pasa como con los mecánicos, es cuestión de promedios, la gran mayoría de lo que se lee en el *New York Times* es más probablemente cierto que lo que se lee, por ejemplo, en el *New York Post*[124].

Las fuentes respetables quieren asegurarse de la veracidad de los hechos antes de publicarlos. En la web pueden encontrarse muchas fuentes que no mantienen esos estándares y en algunos

casos ofrecen noticias sensacionales y rigurosas antes de que los medios más tradicionales y prudentes las ofrezcan. Muchos de nosotros nos enteramos de la muerte de Michael Jackson por TMZ.com, antes de que los medios tradicionales ofrecieran la información. TMZ aceptó publicar el suceso apoyándose en menos pruebas que *Los Angeles Times* o la NBC. En este caso particular, resultó que TMZ estaba en lo cierto, pero no debemos fiarnos de ese tipo de periodismo.

Una cierta cantidad de noticias que han circulado en Twitter sobre la muerte de personas famosas han resultado ser falsas. Sólo en 2015 encontramos a Carlos Santana, James Earl Jones, Charles Manson y Jackie Chan. Un tweet falso en 2011[125] provocó una venta masiva de acciones de la empresa Audience, Inc., y esta perdió el 25% de su valor. La propia Twitter vio cómo sus acciones subían un 8% temporalmente, tras la divulgación en Twitter de unos falsos rumores sobre su compra, los rumores se basaban en una web fraudulenta que se asemejaba mucho a Bloomberg.com. Tal como informó el *Wall Street Journal,* «el empleo de rumores y noticias falsas para manipular el valor de las acciones es una treta que tiene cientos de años. La diferencia con la actualidad es la ubicuidad y la cantidad de información que circula por los mercados y que hace muy difícil que los agentes de bolsa que trabajan a enorme velocidad puedan evitar timos bien diseñados»[126]. Esto le puede ocurrir a los mejores; el veterano reportero (que forma parte del equipo de periodistas que ganó un Premio Pulitzer en 1999) Jonathan Capehart escribió una historia para el *Washington Post* basada en un tuit de un congresista inexistente de un distrito inexistente[127].

Al igual que con los gráficos y las estadísticas, no es deseable que creamos todo lo que nos encontramos que proviene de una fuente fiable, ni hemos de rechazar automáticamente todo lo que tiene su origen en una fuente dudosa. No debemos creer todo lo que leemos en el *New York Times,* ni rechazar todo lo que leemos en TMZ. El origen de una afirmación le aporta credibilidad y, al igual que en un proceso judicial, no debemos apoyarnos en un único testigo, es necesario procurar corroborar las pruebas.

El dominio de las webs

El sufijo de tres letras de la URL indica el dominio. Es importante familiarizarse con los dominios de nuestro país, ya que algunos de ellos tienen restricciones y eso puede ayudarnos a establecer la credibilidad de una web para un tema dado. En los Estados Unidos, por ejemplo, .edu está reservado para instituciones educativas sin ánimo de lucro, como Stanford.edu (la Universidad de Stanford); .gov se reserva para agencias gubernamentales oficiales, como CDC.gov (los Centros de Control de Enfermedades); .mil para las organizaciones militares estadounidenses, como army.mil. El más famoso probablemente es .com, que se utiliza para empresas comerciales, como GeneralMotors.com. Otros incluyen .net, .nyc, y .management, que no suponen ninguna restricción (¡). *Caveat emptor*[128]. BestElectricalService.nyc (ElMejorServicioElectrico.nyc) en realidad podría estar en Nueva Jersey (y sus empleados podrían carecer siquiera de licencia para trabajar en Nueva York); AlphaAndOmegaConsulting.management podría no tener ni idea de procesos de gestión empresarial.

Conocer el dominio también puede ayudarnos a identificar posibles sesgos. Es más probable que encontremos un informe neutro en un estudio educativo o sin ánimo de lucro (en una web .edu, .gov o .org) que en una web comercial, aunque tales webs también pueden alojar blogs de estudiantes y opiniones sin contrastar. Las webs educativas y sin ánimo de lucro no carecen de sesgos: es posible que incluyan información a fin de maximizar las donaciones o mejorar el apoyo público para sus funciones. Pfizer.com puede mostrar sesgos al comentar fármacos elaborados por empresas de la competencia, como GlaxoSmithKline, y desde luego Glaxo puede mostrar sesgos hacia sus propios productos.

Conviene resaltar que no siempre buscamos la neutralidad. Cuando buscamos el manual del usuario de nuestra nevera es probable que visitemos una web favorable al fabricante de la misma (por ej., Frigidaire.com), en lugar de otra que podría estar divulgando versiones obsoletas o erróneas del manual. Una web

.gov puede estar sesgada a favor de los intereses del gobierno, pero es el lugar donde encontrar información más rigurosa sobre leyes, normas fiscales, datos del censo o cómo registrar un automóvil. CDC.gov y NIH.gov[129] probablemente ofrecen información más rigurosa sobre cuestiones médicas, ya que no tienen intereses financieros.

¿Quién está por detrás?[130]

¿Es posible que tenga intención de engañarnos una web que opera bajo un cierto nombre? La Asociación de Productores de Vitamina E podría crear una web denominada NutritionAndYou.info, justo para hacernos creer que sus informaciones no están sesgadas. El presidente de la cadena de tiendas de alimentación Whole Foods (Alimentos integrales) fue pillado haciéndose pasar por un cliente en la red, divulgando la calidad de la comida de su empresa. Muchas webs que incluyen calificaciones, incluyendo Yelp! y Amazon, han encontrado sus buzones de voto atiborrados por amigos y familiares de las personas y productos que se califican. Las personas no siempre son quienes parecen ser en la red. Sólo porque una web se denomina como el Servicio de Salud del Gobierno de los Estados Unidos no significa que sea controlada por el gobierno; que una web se llame Laboratorios Independientes no significa que sea independiente, bien podría estar siendo manejada por un fabricante de automóviles que desea que sus coches salgan bien parados en test no tan independientes.

En la campaña electoral por el distrito trece de Florida, en 2014[131], la oficina local del GOP[132] creó una web con el nombre de su oponente demócrata, Alex Sink, para engañar a la gente haciéndola creer que donaba dinero para ella; en realidad el dinero iba a parar a su oponente, David Jolly. La web, contribute.sinkforcongress2014.com, utilizaba la combinación de colores de Sink e incluía una foto de ella sonriendo, muy similar a la foto de su propia web.



Imagen de la web de la candidata demócrata al Congreso, Alex Sink (Trabajando juntos, con todas las tendencias, para acabar con la parálisis de Washington).

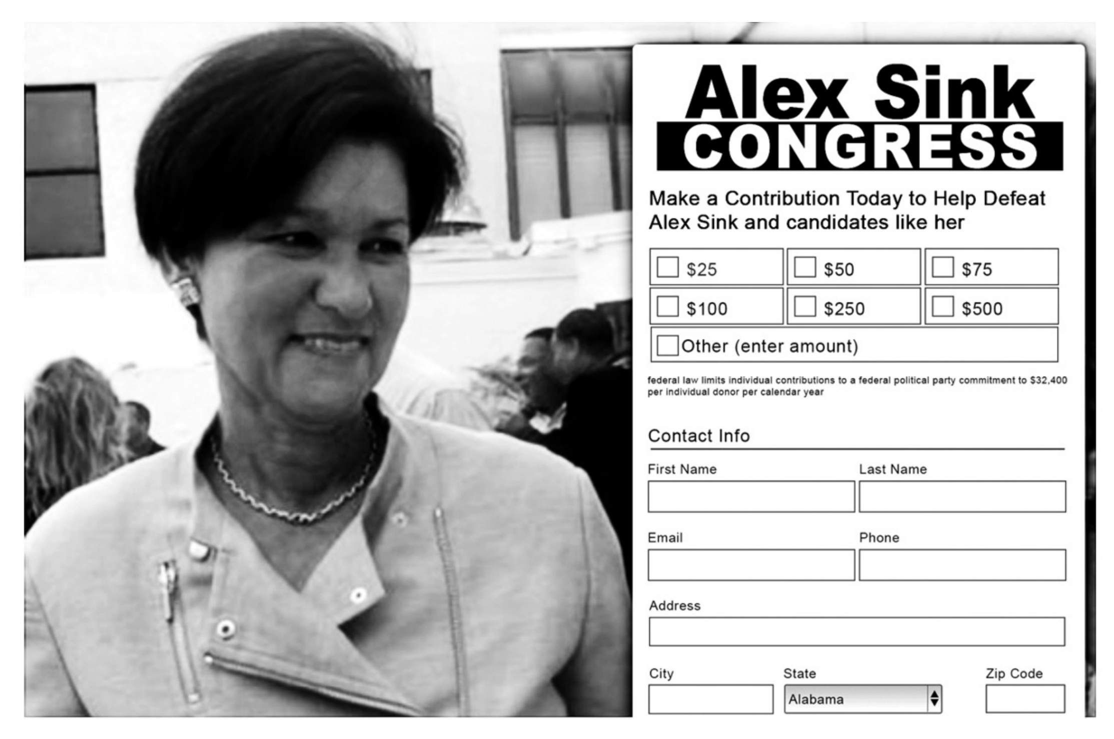


Imagen de la web del GOP utilizada para solicitar dinero para el oponente republicano de Alex Sink, David Jolly
(Haz una aportación ahora para ayudar a derrotar a Alex Sink y a candidatos como ella).

La web del GOP dice que el dinero va a emplearse para derrotar a Sink, no es un fraude declarado, pero convengamos, la mayoría de las personas no se detienen a leer atentamente ese tipo de texto. Las partes más llamativas de la web capciosa son una gran foto de Alex Sink y el encabezamiento «Congreso», lo que sugiere poderosamente que la web es *a favor* de Alex Sink y no contra ella. A fin de no quedarse atrás los demócratas

respondieron con el mismo truco y crearon una web llamada www.JollyForCongress.com para recaudar fondos dirigidos al rival de Sink.

Dentec Safety Specialists y Degil Safety Products son empresas que compiten ofreciendo servicios y productos semejantes. Dentec tiene una web, DentecSafety.com, para comercializar sus productos y Degil tiene una web, DegilSafety.com. Sin embargo, Degil también registró DentecSafety.ca para redirigir a los clientes canadienses a su propia web, para robarles los clientes. Un juez dictaminó que Degil[133] debía pagar a Dentec 10.000 dólares y abandonar DentecSafety.ca.

Un vendedor en línea operaba una web llamada GetCanadaDrugs.com (GetCanadaDrugs.com)[134]. Un tribunal dictaminó que el nombre de la web es «engañosamente mal denominado». Los argumentos principales para ello apuntaban a que no todos los productos farmacéuticos eran fabricados en Canadá y que únicamente en torno al 5% de los clientes de la web eran canadienses. El nombre de ese dominio ha dejado de existir.

Conocer el nombre de un dominio ayuda, pero no constituye un sistema de verificación a prueba de errores. MartinLutherKing.org suena como un lugar donde obtener información sobre el gran orador y líder de los derechos civiles. Como se trata de una web .org, podríamos creer que no tiene ánimo de lucro. La web proclama que ofrece «una valoración histórica auténtica» de Martin Luther King. Pero esperad un momento. La mayor parte de las personas no comienzan a hablar afirmando «lo que voy a contarte es verdad». La BBC no empieza sus noticias diciendo «lo siguiente es verdad». La verdad es la postura por omisión que adoptamos cuando creemos que los demás van a contarnos la verdad. Tal como dice el antiguo refrán: «¿Cómo sabes si alguien te está mintiendo? Porque empieza la frase diciendo *para ser completamente honesto*...». Las personas honestas no necesitan iniciar así sus comentarios.

Lo que se encuentra en MartinLutherKing.org[135] es una serie vergonzante de distorsiones, diatribas antisemitas y citas fuera de contexto. ¿Quién controla esa web? Stormfront, un grupo supremacista blanco y neonazi[136]. ¿Qué mejor manera de ocultar un programa racista que prometiendo «la verdad» sobre un gran líder de los derechos civiles?

Sesgo institucional

¿Hay sesgos que pueden actuar sobre el modo como una persona o una organización estructuran y presentan la información? ¿Tienen esa persona o esa organización un conflicto de intereses? Una declaración sobre el saludable valor de las almendras realizada por la Asociación de Productores de Almendras no resulta tan creíble como una realizada por un laboratorio independiente.

Cuando valoramos a los expertos hay que tener en cuenta que éstos pueden ser sesgados incluso sin ser conscientes de ello. Para un mismo tumor un cirujano puede recomendar la cirugía, en tanto que un radiólogo recomienda la radioterapia y un oncólogo la quimioterapia. Un psiquiatra puede que recomiende fármacos contra la depresión, mientras que un psicólogo recomendaría una terapia verbal. Como dice el dicho, para quien tiene un martillo, todo parece clavo. ¿Quién está en lo cierto? Es probable que necesitemos echar un vistazo a las estadísticas por nuestra cuenta, o buscar a alguien neutral que haya evaluado las distintas posibilidades. Para eso sirven los metaanálisis en la ciencia y la medicina. (O al menos se supone que sirven para eso). El metaanálisis es una técnica de investigación mediante la que los resultados de docenas o centenares de estudios de distintos laboratorios se analizan conjuntamente para determinar la fuerza de las pruebas que apoyan una cierta afirmación. Es la razón por la que las empresas contratan auditores para que analicen su contabilidad o a un analista financiero para decidir si la empresa que se plantean comprar realmente vale lo que dice. Quienes trabajan en la empresa que va a ser comprada sin duda

son expertos sobre la situación financiera de su empresa, pero estarían claramente sesgados. Y no siempre en la dirección que uno pensaría; puede ser que inflen el valor de la empresa si desean venderla, o que lo disminuyan, si están preocupados por una compra hostil.

¿Quién está vinculado a una página web?

Un tipo de búsqueda especial de Google nos permite conocer quién más está vinculado *(link)* a una página web que estamos viendo. Basta con teclear «link:» seguido de la URL de la web y Google nos devuelve todas las web que están vinculadas a ella. (Por ejemplo, link:breastcancer.org nos muestra dos centenas de webs que están vinculadas a ella). ¿Para qué nos sirve hacer esto? Si una agencia de protección del consumidor, Better Business Bureau, o cualquier otra organización de vigilancia, está vinculada a una web, puede interesarnos saber si la elogian o la critican. La página web podría formar parte de un proceso legal, o podría estar vinculada a una fuente de autoridad, como la Sociedad Americana contra el Cáncer, como recurso de interés.

Alexa.com nos informa sobre la demografía de quienes visitan una web, de qué país son, su nivel educativo y qué web han visitado inmediatamente antes de visitar la página en cuestión. Esa información puede proporcionarnos una imagen más clara de quién está utilizando una web y una idea de los motivos por los que lo hace. Una web sobre fármacos que es visitada por médicos probablemente sea una fuente más fiable que otra que no recibe esas visitas. Los comentarios sobre una empresa local realizados por personas de esa misma ciudad probablemente sean más relevantes que los comentarios realizados por personas de fuera de ese Estado.

Revistas con revisión por expertos

En las publicaciones con revisión por expertos, académicos que no tienen relación entre sí evalúan experimentos, informes,

teorías o afirmaciones. Deben ser especialistas en el ámbito que evalúan. El método está lejos de ser infalible y a veces hay descubrimientos que han de revocarse o artículos que se retiran. La revisión por expertos no es el único método existente, pero proporciona una buena base para que podamos obtener conclusiones y, como en una democracia, es el mejor de los sistemas posibles. Si algo es publicado en *Nature, The Lancet* o *Cell,* podemos estar seguros de que ha sido sometido a una revisión rigurosa por expertos. Si nos debatimos a la hora de prestar confianza a un tabloide o a un medio de comunicación serio, todo apunta a que será más correcto lo que ha aparecido publicado en una revista con revisión por expertos.

En una publicación científica o académica la información debe incluir notas a pie de página u otras citas de bibliografía revisada por expertos académicos. Lo que se afirma ha de justificarse, los hechos deben documentarse mediante citas de fuentes respetables. Hace diez años resultaba relativamente sencillo saber si una revista era respetable, pero hoy día la frontera está menos clara por la proliferación de revistas de acceso libre, que publican lo que sea a cambio de una tarifa, en un mundo paralelo pseudoacadémico. Los bibliotecarios especializados en textos académicos pueden ayudarnos a diferenciar unas de otras. Las revistas que aparecen en índices como PubMed (mantenido por la Biblioteca Nacional de Medicina de los Estados Unidos) son seleccionadas por su calidad; los artículos que pueden obtenerse mediante una búsqueda normal, no. Scholar.Google.com es más restrictivo que Google y que otros buscadores, y limita sus resultados a publicaciones científicas y académicas, aunque no supervisa las revistas e incluye muchas publicaciones pseudoacadémicas. Realiza un buen trabajo al expurgar materiales que ni siquiera se *parecen* a la investigación científica, pero es un cuchillo de doble filo: eso puede hacer que resulte más difícil saber qué aceptar entre muchos resultados que parecen ser válidos. Jeffrey Beall, un investigador bibliotecario de la Universidad de Colorado en Denver, ha elaborado una lista negra de las que denomina revistas de acceso libre predadoras (que frecuentemente cobran elevadas tarifas por publicar). Su

lista ha crecido de las veinte publicaciones que incluía hace cuatro años a las más de trescientas que incorpora hoy. Hay otras webs que pueden ayudaros a expurgar publicaciones científicas, como el Social Science Research Network (ssrn.com).

Autoridades reguladoras

En la red no existe una autoridad central que impida a las personas realizar afirmaciones que no son verdad, ni hay otro modo de cerrar una web injuriosa que someterse a un costoso procedimiento legal, hasta obtener una sentencia.

Fuera de la web resulta más fácil hacerse una idea de cómo están las cosas. La exactitud de los libros de texto y enciclopedias es sometida a una rigurosa revisión por expertos (aunque los contenidos a veces cambian por presiones de los comités escolares y los legisladores). Los artículos que se publican en los principales periódicos de los países democráticos contrastan minuciosamente sus fuentes, si se compara con lo que sucede en los periódicos no fiables de Irán o de Corea del Norte, por ejemplo. Si un fabricante de fármacos hace una afirmación, la FDA en los Estados Unidos (el Health Canada en Canadá y otras agencias semejantes en los demás países) han de certificarlo. Si aparece un anuncio en la televisión, el FTC investiga si lo que afirma no es cierto o lleva a la confusión (en Canadá se encarga de eso la ASC, Advertising Standards Canada; en el Reino Unido, la ASA, Advertising Standards Authority; la Unión Europea utiliza una organización de autorregulación denominada EASA, European Advertising Standards Alliance, y muchos otros países disponen de mecanismos equivalentes).

Los mentirosos redomados que realizan afirmaciones fraudulentas pueden enfrentarse a condenas, pero a menudo estas son leves y no resultan disuasorias. El fabricante de bebidas energéticas Red Bull[137] pagó más de 13 millones de dólares en una demanda colectiva por inducir engañosamente a los consumidores a creer que obtendrían un rendimiento físico y mental mejor. En 2015, Target aceptó pagar 3,9 millones de

dólares[138] para que se retirase una demanda por cobrar precios más elevados en las tiendas que lo que se ofrecía en los anuncios que, además, no reflejaban correctamente el peso de sus productos. La empresa alimentaria Whole Foods también fue condenada en 2015 por no reflejar adecuadamente el peso de sus alimentos en los embalajes. Kellogg's pagó 4 millones de dólares[139] para alcanzar un acuerdo en un proceso por haber anunciado fraudulentamente que sus Frosted Mini-Wheats contaban con «pruebas clínicas que demuestran que mejora la atención de los niños un 11%». Pese a que esas cifras nos parezcan abultadas, para Red Bull (con unos ingresos de 7.700 millones de dólares en 2014), Kellogg's (14.600 millones) y Target (72.600 millones) esas multas son poco más que un error de redondeo en su contabilidad.

¿La información está actualizada? ¿Está desacreditada?

Al contrario de lo que sucede con los libros, los periódicos y otras fuentes convencionales, las páginas web raramente indican su fecha; los gráficos, los mapas y las tablas de datos no siempre indican a qué periodo temporal se refieren. No se debe suponer que el «Ventas anuales hasta el momento» que aparece en una página web hoy realmente abarca hoy en «hasta el momento», o siquiera si se refiere a este año.

Dado que las páginas web son relativamente baratas y fáciles de crear, la gente a menudo deja de mantenerlas cuando ya no las necesita, se dedican a otros proyectos y no se sienten obligados a actualizarlas. Se convierten en el equivalente en línea de un escaparate con un letrero de neón que dice «abierto» cuando, en realidad, la tienda está cerrada.

A causa de las distintas razones que he mencionado, fraude, incompetencia, errores de medida, errores de interpretación, los descubrimientos y los inventos pueden acabar desacreditados. Los individuos que han sido considerados culpables tras un juicio legal acaban siendo despedidos. Los airbags sometidos a múltiples inspecciones acaban retirados. Los especialistas

cambian de opinión. Limitarse a considerar la novedad de una web no es suficiente para asegurar que no ha sido desacreditada. Prácticamente cada semana aparecen webs nuevas que afirman cosas que ya han sido plenamente desmentidas. Hay muchas webs dedicadas a desvelar mitos urbanos, como Snopes.com, o a recopilar retractaciones, como RetractationWatch.com.

Durante el otoño de 2015, antes de las elecciones presidenciales de Estados Unidos de 2016, bastantes personas accedieron a webs que verificaban las afirmaciones que hacían los políticos. Los políticos han estado mintiendo al menos desde que Quinto Tulio Cicerón aconsejó a su hermano Marco que lo hiciera en el 64 a.C. Nosotros disponemos de algo que Cicerón no tenía: la verificación en tiempo real. Lo que no significa que todas las verificaciones sean rigurosas o carentes de sesgos, querido lector, continúa siendo necesario asegurarse de que los verificadores carecen de sesgos a favor o en contra de un candidato o de un partido.

Politifact.com es una web que depende del periódico *Tampa Bay Times* que ganó un Premio Pulitzer por sus informaciones dedicadas a supervisar y verificar discursos, apariciones públicas y entrevistas realizadas por figuras políticas y que emplea una escala de seis puntos para calificar las afirmaciones como Verdad, Verdad en su mayor parte, Verdad a medias, Falso en su mayor parte, Falso y, en el extremo de la falsedad, Mentira podrida, para aquellas afirmaciones que no son rigurosas y además son completamente ridículas. El *Washington Post* también mantiene una web de verificación de sucesos[140], con una clasificación que abarca de uno a cuatro Pinochos y que otorga la reconocida medalla de Geppetto a aquellas declaraciones que contienen «la verdad, toda la verdad y nada más que la verdad».

A modo de ejemplo, el candidato presidencial Donald Trump habló en un mitin el 21 de noviembre de 2015, en Birmingham, Alabama. Para fundamentar su propuesta de crear un registro de musulmanes en los Estados Unidos para combatir la amenaza terrorista interna en el país, relató que había visto a «miles y miles» de musulmanes en Jersey City celebrando el

desmoronamiento del World Trade Center el 11 de septiembre de 2001. El reportero de ABC News, George Stephanopoulos, encaró a Trump al día siguiente ante las cámaras y le indicó que la policía de Jersey City negaba que tal cosa hubiera ocurrido. Trump respondió que lo había visto por la televisión, con sus propios ojos, y que estaba perfectamente documentado. Politifact y el *Washington Post* revisaron todas las grabaciones de las transmisiones y programas de noticias de los tres meses posteriores a los atentados y no localizaron nada que pudiera apoyar las afirmaciones de Trump. De hecho, en Paterson, Nueva Jersey, los musulmanes pusieron en la calle principal una pancarta que decía «La comunidad musulmana no apoya el terrorismo». Politifact resumió sus hallazgos[141] indicando que el recuerdo de Trump «se opone a todas las pruebas que hemos podido encontrar. Calificamos esta afirmación de Mentira Podrida». El *Washington Post* lo calificó con cuatro Pinochos.

Durante la misma campaña, Hillary Clinton afirmó que «todos mis abuelos» fueron inmigrantes. Según Politifact (basándose en el censo de los Estados Unidos), tan solo uno de ellos nació en el extranjero[142]; tres de sus cuatro abuelos nacieron en los Estados Unidos.

¿Copiado y pegado, reenviado, editado?

Un modo de hacer creer a la gente que somos verdaderamente cultos es encontrar cosas que suenen cultas en las páginas web de otros y colocarlas en la nuestra. Y, ya que estamos, ¿por qué no incorporar también nuestras propias opiniones controvertidas, que ahora aparecerán envueltas en la erudición de otros, e incrementar las visitas a nuestra web? Si tenemos la idea de promover alguna campaña, siempre podemos poner a caer de un burro los argumentos cuidadosamente expuestos de otro, los editamos y los empleamos para apoyar posiciones contrarias a las suyas. A los demás nos toca averiguar si realmente es original lo que estamos leyendo, si se trata de información sin contaminar y no de un batiburrillo montado por otro.

Información bien sustentada

Los charlatanes sin escrúpulos se valen de que la mayoría de las personas no se preocupan de leer las notas a pie de página ni rastrean las citas. Eso hace que mentir sea muy fácil. Quizá tenemos interés en que nuestra página web convenza a la gente de que nuestra crema corporal invierte en diez años el proceso de envejecimiento. Así que escribimos un artículo y lo salpimentamos con notas que enlazan con otras páginas web que son completamente irrelevantes para el asunto. Eso podría engañar a muchas personas, ya que la gran mayoría de ellas no consultarán los enlaces y quienes lo hagan probablemente se van a limitar a comprobar que la URL pertenece a una web relevante, aunque el artículo que se cita no tenga nada que ver con nuestro producto.

Se puede hacer algo más diabólico: la cita podría tener una relación marginal, pero no ser relevante. Podríamos afirmar que nuestra crema contiene vitamina X y que se ha demostrado que la vitamina X mejora la salud y la calidad de la piel. Muy bien, pero ¿cómo? Los estudios sobre la vitamina X, ¿se realizaron con personas que se untaron la vitamina o que la tomaron por vía oral? ¿Cuál fue la dosis? ¿La crema que vendemos lleva una cantidad adecuada de vitamina X?

Las trampas de la terminología

Puede que en CDC.gov leamos que la incidencia de una cierta enfermedad es de 1 entre 10.000 personas. Pero poco después nos topamos con un artículo en NIH.gov que afirma que la misma enfermedad tiene una prevalencia de 1 entre 1.000. ¿Es que alguien se confundió al poner el punto, es una errata? ¿La incidencia y la prevalencia no son lo mismo? Pues no, no lo son. La incidencia de una enfermedad es la cantidad de casos nuevos (incidentes) que se conocen en un cierto periodo de tiempo, por ejemplo, un año. La prevalencia es la cantidad de casos que existen, el número total de personas que padecen la enfermedad. (A veces quienes tienen miedo de las cifras comenten el error de

creer a primera vista que 1 entre 1.000 es menos que 1 entre 10.000, atendiendo a la cifra con más ceros, en lugar de a la palabra *entre*).

Consideremos la esclerosis múltiple (EM), una enfermedad que produce la desmielinización del cerebro y la médula espinal. Cada año se diagnostican unos 10.400 casos nuevos en los Estados Unidos, lo que implica una incidencia de 10.400/322.000.000[143], o 3,2 casos por cada 100.000 personas, en otras palabras, hay una probabilidad de 0,0032 de contraer la enfermedad. Comparemos esa cifra con la cantidad de personas que ya padecen la enfermedad en Estados Unidos, 400.000, lo que supone una tasa de prevalencia de 400.000/322.000.000, o 120 casos por cada 100.000 habitantes, una probabilidad de 0,12 de contraer la enfermedad en algún momento de nuestras vidas.

Además de la incidencia y la prevalencia, frecuentemente se cita también una tercera estadística, la tasa de mortalidad, la cantidad de personas que mueren a causa de una enfermedad, habitualmente durante un cierto periodo de tiempo. Cada año se diagnostican 1,1 millón de casos nuevos de la enfermedad de las arterias coronarias[144], en la actualidad hay 15,5 millones de estadounidenses que la padecen y 375.295 mueren por su causa todos los años. La probabilidad de que nos diagnostiquen una enfermedad cardíaca este año es del 0,3%, aproximadamente cien veces más que la de padecer EM; la probabilidad de que la padezcamos ahora mismo es casi del 0,5% y la de morir por su causa en cualquier año es del 0,1%. La de morir a causa de esta enfermedad en un momento dado de nuestras vidas es del 20%. Como vimos en la primera parte, estos datos se aplican a la suma de todos los estadounidenses. Si conocemos una persona concreta, su historial familiar de enfermedades cardíacas, si fuma o no, su peso y su edad, podemos realizar estimaciones más afinadas, empleando la probabilidad condicional.

La tasa de incidencia de una enfermedad puede ser elevada, al tiempo que sus tasas de prevalencia y de mortalidad sean relativamente bajas. El catarro común es un ejemplo: hay muchos millones de personas que se acatarran en un año (incidencia elevada), pero en casi todos los casos se cura rápidamente y, por

lo tanto, la prevalencia, la cantidad de personas que lo padecen en un momento dado, puede ser baja. Algunas enfermedades son relativamente raras, crónicas y fáciles de curar, de modo que su incidencia es baja (no hay muchos casos al año), pero su prevalencia es elevada (se suman todos los casos y las personas continúan viviendo con la enfermedad) y la mortalidad es baja.

Cuando las personas evalúan las pruebas a menudo no prestan atención a las cifras y las denominaciones de los ejes, como hemos visto, pero además también desconocen los descriptores verbales. Recordad los mapas de Estados Unidos que mostraban la tasa bruta de nacimientos en la primera parte. ¿Os preguntasteis qué es la «tasa bruta de nacimientos»? Podemos imaginar que una tasa de nacimientos puede ajustarse según diversos criterios, como si el recién nacido está vivo o no, si el bebé sobrevive durante un cierto periodo de tiempo, etc. Podemos creer que como la definición del diccionario indica que «bruto» es algo en estado natural, no procesado o refinado (por ejemplo un diamante en bruto), eso significa que se trata de cifras no elaboradas, sin adulterar, sin ajustes. Pero no es así. Los estadísticos emplean el término tasa bruta de nacimientos para contabilizar los nacidos vivos (por tanto es una cifra ajustada que elimina los mortinatos). Si vamos a abrir una empresa de pañales, lo que necesitamos es la tasa bruta de nacimientos, no la tasa total (ya que la tasa total incluye los bebés que no sobreviven al nacimiento).

Por cierto, hay una estadística relacionada con la anterior, la tasa bruta de muertes, y se refiere a la cantidad de personas que mueren a cualquier edad. Si restamos esa tasa de la tasa bruta de nacimientos, obtenemos una estadística por la que se interesan mucho quienes desarrollan políticas públicas (a Thomas Malthus le interesaba): el INP, la tasa de incremento natural de una población.

---

120 S. M. Gartler (2006), «The chromosome number in humans: a brief history», *Nature Reviews Genetics,* 7, 655-660, http://www.nature.com/scitable/content/The-chromosome-number-in-humans-a-

brief-15575. B. Glass (1990), *Theophilus Shickel Painter,* Washington, D.C., National Academy of Sciences, http://www.nasonline.org/publications/biographical-memoirs/memoir-pdfs/painter-theophilus-shickel.pdf, consultado el 6 de noviembre de 2015.

121 Paul Simon, Stevie Wonder y Joni Mitchell pueden considerarse expertos en componer canciones. Pese a que no ostenten puestos en la universidad, expertos universitarios han escrito libros y artículos sobre ellos y el señor Simon y el señor Wonder han recibido el reconocimiento del presidente de los Estados Unidos con los premios del Kennedy Center, reservados a individuos que han realizado grandes contribuciones a las artes interpretativas. La señora Mitchell ha recibido un doctorado honorario en música y ganó el premio Polaris Music.

122 *N. del T.:* Se conoce como efecto Mozart a los supuestos y nunca demostrados efectos benéficos que proporcionaría escuchar música de este compositor.

123 N. Chomsky (25 de mayo de 2015), «The *New York Times* is pure propaganda», *Salon,* http://www.salon.com/2015/05/25/noam_chomsky_the_new_york_times_is_pure_proganda_partner/.

M. Achbar, A. Symansky y P. Wintonick (prods.) y M. Achbar, A. Symansky y P. Wintonick (dirs.) (1992), *Manufacturing Consent: Noam Chomsky and the Media* (película), Estados Unidos, Buylines.com Inc. Y Seitgeist Films. https://www.youtube.com/watch?v=BsiBl2CaDFg.

124 *N. del T.:* El *New York Times* tiene fama de ser un periódico serio, en tanto que el *New York Post* es un ejemplo de lo que podría considerarse prensa sensacionalista.

125 E. D. Melendez (1 de febrero de 2013), «Twitter stock market hoax drws attention of regulators», http://www.huffingtonpost.com/2013/02/01/twitter-stock-market-hoax_n_2601753.html; http://www.forbes.com/forbes/welcome/.

126 M. Farrell (14 de julio de 2015), «Twitter shares hit by takeover hoax», *Wall Street Journal,* http://www.wsj.com/articles/twitter-shares-hit-by-takeover-hoax-1436918683.

127 7 de septiembre de 2010, «*Washington Post* writer falls for fake congressman Twitter account», *Huffington Post,* actualizado el 7 de septiembre de 2010, http://www.huffingtonpost.com/2010/09/07/washington-post-writer-fa_n_707132.html; http://voices.washingtonpost.com/postpartisan/2010/09/obama_deficits_and_the_ditch.html.

128 *N. del T.:* En latín en el original, «caveat emptor» significa «cuidado por parte del comprador» o «el comprador asume el riesgo».

129 *N. del T.:* NIH (National Institutes of Health, Institutos Nacionales de la Salud) son los principales centros de investigación médica pública de los Estados Unidos.

130 Está tomado literalmente de *The Organized Mind (La mente organizada)*. D. J. Levitin (2014), *The Organized Mind,* Nueva York, Dutton.

131 A. Leary (4 de febrero de 2014), «Misleading GOP website took donation meant for Alex Sink», *Tampa Bay Times,* http://www.tampabay.com/news/politics/stateroundup/misleading-gop-website-took-donation-meant-for-alex-sink/2164138.

132 *N. del T.:* GOP (Grand Old Party), nombre por el que también se conoce al Partido Republicano en Estados Unidos.

133 D. Pink (2013), «Deceiving domain names not allowed», Wickwire Holm, http://www.wickwireholm.com/Portals/0/newsletter/BLU%20Newsletter%20-%20January%202013%20-%20Deceiving%20Domain%20Names%20Not%20Allowed.pdf; S. Bonni (24 de junio de 2014), «The tort of domain name passing off», *Charity Law Bulletin,* 342, Carters Professional Corporation, http://www.carters.ca/pub/bulletin/charity/2014/chylb342.htm.

134 https://www.canadadrugs.com/; https://www.getcanadadrugs.com/ (ya no está disponible); M. Naud (n.d.), «Registered trade-mark canadadrugs.com found deceptively misdescriptive», ROBIC, http://www.robic.ca/admin/pdf/682/293.045E-MNA2007.pdf.

135 La cita incendiaria de la web tiene su origen en un libro de Taylor Branch, *Pillar of Fire (Pilar de fuego),* pero el autor señala que él mismo no escuchó las cintas, sino que fue informado al respecto por tres agentes del FBI que se lo contaron.

136 Entre las fuentes que identifican Stormfront como la «web de odio más importante» de Internet se encuentran:

B. Levin (2003), «Cyberhate: A legal and historical analysis of extremists' use of computer networks in America», en B. Perry (ed.), *Hate and Bias Crime: A Reader,* Nueva York, Routledge, p. 363.

N. Ryan (2004), *Into a World of Hate: A Journey Among the Extreme Right,* Nueva York, Routledge, p. 80.

S. Samuels (1997), «Is the Holocaust unique?», en Alan S. Rosenbaum (ed.), *Is the Holocaust Unique?: Perspectives on Comparative Genocide,* Boulder, CO, Westview Press, p. 218.

G. Bolaffi *et al.* (eds.) (2002), *Dictionary of Race, Ethnicity and Culture,* Thoousand Oaks, CA, Sage Publications, p. 254.

137 L. O'Reilly (8 de octubre de 2014), «Red Bull will pay $10 to customers disappointed the drink didn't actually give them "wings"»,

http://www.businessinsider.com/red-bull-settles-false-advertising-lawsuit-for-13-million-2014-10.

138 Associated Press (11 de febrero de 2015), «Target agrees to pay $3.9 million in false-advertising lawsuit», http://journalrecord.com/2015/02/11/target-agrees-to-pay-3-9-million-in-false-advertising-lawsuit-law/.

139 Federal Trade Commission (20 de abril de 2009), «Kellogg settles FTC charges that ads for Frosted Mini-Wheats were false [Press release]», https://www.ftc.gov/news-events/press-releases/2009/04/kellogg-settles-ftc-charges-ads-frosted-mini-wheats-were-false.

140 https://www.washingtonpost.com/news/fact-checker/.

141 L. Carroll (22 de noviembre de 2015), «Factchecking Trump's claim that thousands in New Jersey cheered when World Trade Center tumbled», http://www.politifact.com/truth-o-meter/statements/2015/nov/22/donald-trump/fact-checking-trumps-claim-thousands-new-jersey-ch/.

142 K. Sanders (16 de abril de 2015), «In Iowa, Hillary Clinton claims "all my grandparents" came to the U.S. from foreign countries», http://www.politifact.com/truth-o-meter/statements/2015/apr/16/hillary-clinton/hillary-clinton-flubs-familys-immigration-history-/.

143 La población de Estados Unidos en el momento en que escribo esto. http//www.census.gov/popclock/.

144 American Heart Association (2015), «AHA Statistical Update», *Circulation,* 131, pp. 434-441. Agradezco a los bibliotecarios de la McGill University, Robin Canuel y Genevieve Gore, por ayudarme a encontrar esta estadística.

EXPLICACIONES IGNORADAS, INFRAVALORADAS Y ALTERNATIVAS

Cuando evaluamos una afirmación o un argumento no preguntamos si hay alguna otra razón, distinta de la que se ofrece, que podría explicar los hechos u observaciones de los que se informa. Siempre hay explicaciones alternativas, lo que hemos de hacer es contrastarlas con la que se nos ha ofrecido y determinar si quien saca la conclusión ha alcanzado la más evidente o probable.

Por ejemplo, nos cruzamos con un amigo en el vestíbulo y no nos devuelve el saludo, podríamos concluir que está enfadado con nosotros. Pero hay explicaciones alternativas, quizá no nos ha visto, o llegaba tarde a una reunión, o estaba preocupado, o realizaba un experimento psicológico, o había hecho un voto de silencio durante una hora, o había sido temporalmente abducido por los ladrones de cuerpos (o tal vez permanentemente abducido).

Las explicaciones alternativas son muy comunes en las pseudociencias y el contraconocimiento y también en la verdadera ciencia. Los físicos que trabajan en el CERN informaron que habían descubierto neutrinos que viajaban a una velocidad más alta que la luz. Tal cosa habría puesto fin a un siglo de teoría einsteiniana. Al final resultó que se trataba simplemente de un cable suelto en el acelerador lineal que provocaba un error de medida. Lo anterior resalta la idea de que un fallo metodológico en un experimento muy complejo es casi siempre la explicación más probable de un suceso cuya existencia nos obligaría a reconstruir nuestra comprensión de la naturaleza del universo.

De modo semejante, si una página web cita experimentos con una mezcla de vitaminas nueva, desconocida hasta ahora, que

dispara nuestro CI veinte puntos y que ¡las empresas farmacéuticas no quieren que lo sepamos!, deberíamos preguntarnos hasta qué punto es probable que nadie lo sepa y si no cabe la explicación alternativa de que hay alguien intentando ganar dinero.

Los mentalistas, los adivinos y los videntes ganan mucho dinero realizando hazañas aparentemente imposibles de lectura de mente. Una de las explicaciones es que han dado con una fuerza secreta y oculta que es contraria a todo lo que conocemos sobre causas y efectos y la naturaleza del espacio-tiempo. Una explicación alternativa es que son ilusionistas que utilizan trucos de ilusionismo y que se limitan a mentir sobre cómo hacen lo que hacen. En apoyo de lo anterior hay que señalar la existencia de ilusionistas profesionales, como James Randi, quien hasta ahora ha logrado utilizar trucos inteligentes para reproducir todas las hazañas que haya realizado un mentalista. A menudo los ilusionistas explican cómo han realizado un truco, a fin de desacreditar a los autoproclamados videntes. Para ser justos imagino que también es posible que sean los *ilusionistas* quienes intentan engañarnos, que son videntes verdaderos que tienen miedo de revelar sus dotes (por temor a ser explotados, secuestrados, etc.) y que lo único que hacen es fingir que están utilizando trucos habilidosos. Aunque, pensándolo bien, consideremos las dos posibilidades: una de ellas nos obliga a deshacernos de todo lo que conocemos sobre la naturaleza y la ciencia y la otra no. Cualquier psicólogo, cualquier agente de la ley, empresario, cónyuge divorciado, diplomático, espía o abogado os dirán que las personas mienten; que lo hacen por una diversidad de razones y que a veces lo hacen con una frecuencia y entusiasmo alarmantes. En fin, si nos enfrentamos a una afirmación que parece improbable, la explicación (alternativa) más probable es que la persona que nos lo cuenta esté mintiendo de un modo u otro.

Las personas que intentan predecir el futuro sin emplear poderes psíquicos, dirigentes militares, economistas, planificadores empresariales, con frecuencia se equivocan en sus predicciones por no tomar en cuenta las explicaciones

alternativas. Esto ha dado lugar a una práctica empresarial denominada planificación de escenarios, tener en cuenta todos los resultados posibles, incluso aquellos que parecen improbables, lo que puede resultar una tarea muy difícil y en la que incluso los expertos pueden fracasar. En 1968 Will y Ariel Durant escribieron[145]:

> En los Estados Unidos la inferior tasa de natalidad de los anglosajones ha disminuido su poder económico y político, en tanto que la superior tasa de natalidad de las familias católicas indica que hacia el año 2000 la Iglesia católica será la fuerza dominante en los gobiernos municipales, estatales y nacional.

Lo que no lograron tener en cuenta fue que durante los treinta y dos años siguientes muchos católicos abandonaron su Iglesia y muchos de ellos también utilizaron los métodos de control de la natalidad, a pesar de las prohibiciones de su Iglesia. Los escenarios alternativos a su concepción en 1968 eran difíciles de imaginar.

Las predicciones sociales y artísticas también pueden dar un vuelco: los expertos opinaban en la época en la que surgieron los Beatles que «las bandas con guitarras están de capa caída». Algunos de quienes escribieron críticas sobre el estreno de la Quinta Sinfonía de Beethoven opinaron negativamente que nadie querría siquiera volver a escucharla. También la ciencia se equivoca; los expertos dijeron que los trenes de alta velocidad nunca serían posibles, ya que los pasajeros morirían asfixiados. Los especialistas creían que la luz se movía a través de un «éter» invisible. La ciencia y la vida no son estáticas. Lo que podemos hacer es sopesar las pruebas y valorarlas por nuestra cuenta, empleando las mejores herramientas disponibles. Una de esas herramientas es el pensamiento creativo para imaginar alternativas a las concepciones que hemos utilizado hasta el momento.

Las explicaciones alternativas a menudo resultan claves para dilucidar los procesos penales. Los efectos del contexto que consideramos en la primera parte y la incapacidad de comprender

que la probabilidad condicional no funciona retroactivamente han provocado la condena de muchos inocentes.

El auténtico razonamiento científico supone establecer dos (o más) hipótesis y proponer las probabilidades de ambas. En un juicio los abogados no debieran centrarse en las probabilidades de que algo coincida, sino en la probabilidad de dos escenarios posibles: ¿cuál es la probabilidad de que las muestras de sangre tengan el mismo origen, frente a la probabilidad de que no sea así? Más exactamente, lo que debiera hacerse es comparar la probabilidad de que coincidan las muestras, suponiendo que el acusado sea culpable, con la probabilidad de que coincidan, suponiendo que sea inocente. También se podría comparar la probabilidad de que el acusado sea inocente, dados los datos, con la probabilidad de que sea culpable, dados los datos. Asimismo necesitamos conocer la exactitud de las medidas. En el año 2015 el FBI anunció[146] que los análisis microscópicos del cabellos eran inexactos el 90% de las veces. Sin ese tipo de prueba resulta imposible tomar una decisión justa e imparcial[147]. Es decir, si estuviéramos hablando tan solo en términos de coincidencia de las muestras, estaríamos considerando únicamente pruebas unilaterales, la probabilidad de que coincidan dada la hipótesis de que el criminal estuviera en la escena del crimen. Lo que no sabemos es la probabilidad de que las muestras coincidan dadas otras hipótesis alternativas y lo que necesitamos es poder comparar esas probabilidades.

Lo anterior sucede constantemente; en un caso juzgado en Gran Bretaña[148], el sospechoso, Dennis Adams, fue acusado únicamente a partir de pruebas de ADN. La víctima no pudo identificarlo en una ronda de reconocimiento y en el juicio declaró que Adams no se parecía al asaltante. La víctima añadió que Adams parecía unos veinte años mayor que el asaltante. Además, Adams tenía una coartada para la noche del crimen que fue corroborada por una tercera persona. La única prueba que presentó la acusación fue la coincidencia del ADN. Pues bien, Adams tenía un hermano con quien también hubiera debido compararse el ADN, pero no había ninguna prueba adicional de

que el hermano hubiera cometido el crimen, de modo que los investigadores no lo tuvieron en cuenta. Tampoco tenían más pruebas contra Dennis, la *única* prueba era la coincidencia del ADN. Nadie en la sala tuvo en cuenta la hipótesis alternativa de que podía haber sido el hermano de Dennis... Dennis fue condenado tanto en primera instancia como en la apelación.

Construido por los antiguos para que se viera desde el espacio

Posiblemente habréis oído la especulación según la cual la vida humana no evolucionó realmente en la Tierra, sino que una raza de extraterrestres vino y sembró aquí las semillas de la vida humana. En sí mismo no es inverosímil, la cuestión es que no hay ninguna prueba real que apoye tal cosa. Eso no significa que no sea verdad y tampoco significa que no debamos buscar esas pruebas, pero el hecho real es que algo que *pudiera* ser verdad tiene una utilidad limitada, excepto tal vez en la ciencia ficción.

En el año 2015 una noticia aparecida en el *New York Times* describía una misteriosa formación en el suelo de Kazajistán que únicamente podía observarse desde el espacio[149].

> Imágenes tomadas por satélite de una remota y árida estepa en el norte revelan colosales movimientos de tierras, figuras geométricas de cuadrados, cruces, líneas y anillos del tamaño de varios campos de fútbol, reconocibles únicamente desde el aire y los más antiguos de los cuales datan de hace ocho mil años.
>
> El mayor de todos, un asentamiento neolítico, es un gigantesco cuadrado formado por 101 montículos, cuyas esquinas están conectadas por una cruz diagonal que abarca más terreno que la Gran Pirámide de Keops. Otro es una especie de esvástica de tres brazos, los cuales terminan en zigzag doblados en sentido contrario a las agujas del reloj.

Resulta fácil dejarse llevar por la imaginación y pensar que esos enormes dibujos fueron un modo de los antiguos de enviar señales a los extraterrestres, quién sabe si siguiendo unas estrictas instrucciones alienígenas. Tal vez fuera un aeropuerto

para naves espaciales, o un mensaje codificado, algo como «enviad más comida». Los humanos somos así, nos encanta imaginar cosas extraordinarias. Somos una especie dada a la narrativa.

Dejando de lado el hecho bastante obvio de que una civilización capaz de realizar vuelos interestelares sin duda dispondría de medios de comunicación más eficaces que levantar montículos de piedras sobre el terreno, hay una explicación alternativa. Afortunadamente el *New York Times* nos la proporciona (aunque no fuera ese el caso de otros medios que reprodujeron la noticia), mediante una cita de Dimitriy Dey, el descubridor de las piedras misteriosas:

> «No creo que se construyeran para ser vistas desde el espacio», declaró el señor Dey, de 44 años, en una entrevista en su ciudad natal, Kostanay, desdeñando las especulaciones extravagantes sobre los extraterrestres y los nazis. (Mucho antes de Hitler, la esvástica era un elemento de diseño muy antiguo y ubicuo). Su teoría es que las figuras fueron construidas siguiendo líneas rectas sobre elevaciones para «servir como observatorios para seguir el movimiento del sol naciente».

Una explicación basada en el reloj de sol resulta más probable que los alienígenas del espacio. No significa que sea cierta, pero una parte de la capacidad de analizar la información disponible y valorar las afirmaciones que se nos proponen supone desvelar alternativas verosímiles, como esta.

El grupo de control ausente

El denominado efecto Mozart quedó desacreditado porque los experimentos que indicaban que escuchar a Mozart veinte minutos al día eleva el CI carecían de grupo de control. Es decir, a un grupo de personas se les hacía escuchar a Mozart y a otro grupo de personas no se les hacía escuchar nada. No hacer nada no es un control adecuado para hacer algo y resulta que si se proporciona a las personas algo que hacer, prácticamente cualquier cosa, el efecto desaparece. El efecto Mozart no era

provocado porque escuchar a Mozart aumente el CI, sino porque el aburrimiento de no hacer nada temporalmente disminuye el CI.

Si llevamos a veinte personas con dolor de cabeza a un laboratorio, les proporcionamos un nuevo fármaco milagroso contra el dolor de cabeza y diez de ellos mejoran, no habremos aprendido nada. Algunos dolores de cabeza remiten solos. ¿Cuántos? No lo sabemos. Para eso necesitaríamos disponer de un grupo de control formado por personas con edades e historiales semejantes y que padezcan dolores semejantes. Dado que la mera creencia de que podemos curarnos provoca una mejora de la salud, es necesario proporcionar al grupo de control algo que les permita desarrollar esa creencia en la misma medida que al grupo al que se administra la medicina que queremos estudiar. Es el conocido placebo, una pastilla que parece exactamente igual que el fármaco milagroso contra el dolor de cabeza, de modo que nadie sabe quién está tomando qué hasta que termina el experimento.

Malcolm Gladwell divulgó una conclusión no válida en su libro *David and Goliath* al sugerir que las personas con dislexia en realidad podrían disfrutar de una ventaja en sus vidas, lo que condujo a muchos padres a creer que sus hijos disléxicos no debían recibir el apoyo educativo que necesitaban. Gladwell se equivocó por carecer de una condición de control. Desconocemos cuánto *más* éxito habrían tenido los disléxicos que eligió, si hubieran podido mejorar de su trastorno.

El grupo de control ausente sale a relucir en las conversaciones de todos los días, en las que resulta más difícil encontrarlo que en la ciencia, por la simple razón de que no lo buscamos. Leemos y aceptamos una investigación nueva que demuestra que acostarse por la noche y levantarse por la mañana mejora simultáneamente la productividad y la creatividad. Una artista amiga nuestra, con éxito desde cualquier punto de vista, contesta que siempre durmió cuando quiso, que a menudo se ha pasado noches en vela y que a veces duerme veinte horas seguidas y que le va la mar de bien. Pero falta un grupo de control. ¿Cuánto *más* productiva y creativa habría sido, si hubiera dormido regularmente?[150] No lo sabemos.

Dos gemelos fueron separados al nacer y criados separados, uno en la Alemania nazi y el otro en Trinidad y en Venezuela[151]. Uno fue educado como católico e ingresó en las juventudes hitlerianas, el otro fue educado como judío. Se reunieron veintiún años después[152] y descubrieron una sorprendente lista de conductas semejantes que muchas personas fascinadas atribuyeron a la genética: ambos gemelos se rascaban la cabeza con el dedo anular, ambos se divertían acercándose con sigilo a los extraños y estornudaban ruidosamente. Los dos hombres tenían bigotes pequeños, bien recortados y llevaban gafas cuadradas de alambre, redondeadas en las esquinas. Ambos llevaban camisas azules con hombreras y bolsillos de estilo militar. Los dos compartían la misma forma de andar y se sentaban igual en las sillas. A los dos les encantaba la mantequilla y la comida especiada, tiraban de la cadena del retrete antes y después de usarlo y leían en primer lugar el final de los libros. Ambos envolvían los lápices y las plumas con cinta adhesiva para sujetarlas mejor.

Historias como la anterior pueden hacer que nos preguntemos hasta qué punto nuestra conducta está influida por la genética. O si somos meros autómatas y nuestros actos están predeterminados. ¿Cómo explicar si no tales coincidencias?

Bueno, hay dos modos y ambos conducen al grupo de control ausente. Un psicólogo social podría decirnos que las personas que se parecen son tratadas de modo semejante. Quienes son atractivos son tratados de un modo distinto que quienes no lo son, los altos reciben un tratamiento diferente que los bajos. Si nuestro rostro sugiere que somos honestos y desprendidos, las personas nos tratarán de un modo diferente que si nuestra cara sugiere algo distinto. Las conductas de los hermanos fueron moldeadas por el mundo social en el que vivían. Se necesita un grupo de control de personas sin relación, pero que se parezcan mucho y que fueran criadas separadas, para poder extraer alguna conclusión clara de ese «experimento natural» de los gemelos separados al nacer.

Un estadístico o un genetista de la conducta[153] nos dirían que entre las miles y miles de cosas que hacemos, hay una probabilidad de que dos extraños compartan algunas semejanzas sorprendentes al vestirse, acicalarse, la afición por las bromas o alguna inclinación rara, son cosas que podríamos encontrar si las buscáramos con atención y durante el tiempo necesario. Sin ese grupo de control, reunir extraños y hacer un inventario de sus costumbres, no podemos saber si la fascinante historia de los gemelos se debe a la genética o al mero azar. Es posible que la genética tenga un papel, pero probablemente no es un papel tan importante como podríamos creer.

Selección interesada

Nuestros cerebros están conformados para elaborar historias a medida que asimilan la enorme amplitud del mundo en el que suceden millones de eventos cada segundo. Es fácil que ocurran coincidencias que realmente no tienen significado alguno. Si una amiga a la que hace tiempo que no vemos nos llama cuando estamos pensando en ella, no significa que ninguno de los dos tenga poderes psíquicos. Si ganamos tres veces seguidas a la ruleta, no significa que estemos en racha y que debamos apostar hasta nuestro último dólar en la próxima jugada. Si nuestro mecánico no cualificado nos arregla el coche, no significa que será capaz de hacerlo la próxima vez, es posible que tuviera suerte.

Puede que tengamos una teoría favorita, como que un exceso de vitamina D causa malestar y es posible que encontremos pruebas que la apoyen. Pero si nos limitamos a buscar pruebas que apoyen la teoría, no estaremos haciendo investigación de modo correcto, ya que estaríamos ignorando las pruebas en contra, puede que haya muchas o pocas, pero no lo sabremos, ya que no las hemos buscado. En términos coloquiales los científicos denominan ese proceso «selección interesada» de los datos que coinciden con nuestra teoría. La verdadera investigación exige que mantengamos abierta la mente y que nos

atrevamos a tener en cuenta tanto las pruebas favorables como las contrarias, para poder establecer una conclusión basada en las pruebas (y no una conclusión basada en «caramba, me gustaría que las cosas fueran así»).

La selección interesada suele ir acompañada del foco selectivo. Este tiene lugar cuando la información a la que accedemos no representa al conjunto. Si observamos una ciudad desde la ventana de un tren, tan solo vemos una parte de ella y no se trata necesariamente de una parte representativa, únicamente tenemos acceso visual a la parte de la ciudad por la que pasan la vías del tren, con todos los sesgos que eso conlleva. Los trenes hacen ruido, las personas con más recursos habitualmente viven en casas alejadas del ruido, de modo que quienes viven cerca de las vías del tren suelen ser los de ingresos más bajos. Si lo que conocemos de la ciudad se limita a quienes tienen ingresos más bajos, no conocemos toda la ciudad.

Lo anterior tiene relación con los comentarios de la primera parte sobre la recopilación de datos (cómo se recogen los datos) y la importancia de obtener muestras representativas. Pretendemos comprender la naturaleza del mundo, o al menos de la ciudad por la que pasa el tren, y hemos de tener en cuenta explicaciones alternativas de lo que vemos o de lo que se nos dice. Una buena explicación alternativa con aplicaciones muy diversas es que solo vemos una parte del cuadro y la parte que no estamos viendo podría ser muy distinta.

Puede que nuestra hermana nos muestre orgullosa los dibujos de su hija de cinco años. ¡A lo mejor son excelentes! ¡Si os gusta la pintura, enmarcadlos! Pero si os estáis planteando invertir en el futuro de la niña por si acaba siendo la mejor pintora del mundo, tal vez convenga hacerse algunas preguntas: ¿Quién ha recogido el dibujo? ¿Quién lo ha seleccionado? ¿De qué tamaño era el original? ¿Cuántos dibujos hizo el Picasso infantil antes de hacer este? ¿Qué sucedió antes y qué sucedió después? A causa del foco selectivo podríamos estar contemplando una muestra de una serie de dibujos maravillosos o una pequeña joyita que forma parte de una obra mucho mayor (y nada llamativa), que fue identificada y recogida por el profesor.

También observamos el foco selectivo en los titulares de las noticias. Un titular puede anunciar que «El triple de estadounidenses apoyan esta legislación frente a quienes la rechazan». Incluso si, basándonos en los pasos descritos en la primera parte, nos tranquiliza saber que la encuesta fue realizada sobre una muestra representativa y suficientemente numerosa de estadounidenses, no es posible concluir que la mayoría de ellos apoyen la legislación. Sería perfectamente factible que el 1% la rechacen, el 3% la apoye y el 96% estén indecisos. Trasladad este tipo de jugarreta al titular sobre unas elecciones en el que se afirme que cinco veces más republicanos apoyan al candidato A que al candidato B en las primarias para presidente. Podría ser cierto, pero el titular podría estar ocultando que el candidato C tiene un 80% de la intención de voto.

Intentad lanzar una moneda al aire diez veces. «Sabemos» que debería salir cara la mitad de las veces. Pero probablemente no sucederá así. Incluso si la lanzáramos 1.000 veces, probablemente no obtendríamos exactamente 500 caras. La probabilidad teórica únicamente se alcanza mediante una cantidad infinita de ensayos. Cuantas más veces lancemos la moneda, más nos aproximaremos a una proporción 50/50 de caras y cruces. Resulta contrario a la intuición, pero hay una probabilidad muy cercana al 100 por cien de que en algún punto de la secuencia obtengamos cinco caras seguidas. ¿Por qué es tan contrario a nuestra intuición? Nuestro cerebro no ha evolucionado con una comprensión suficiente del azar. Lo que tenderá a suceder no es una secuencia cara-cruz-cara-cruz, sino series (también llamadas plenos), incluso en una secuencia aleatoria. Esto facilita engañar a la gente. Grabaos con el móvil lanzando 1.000 veces seguidas una moneda. Antes de cada lanzamiento decís «esta va a ser la primera de una serie de cinco caras seguidas». Si sacáis una cara, antes del siguiente lanzamiento decís «esta va a ser la segunda cara de una serie de cinco caras seguidas»; si lo que sale es cruz, comenzáis de nuevo. Pero si no lo es, antes de lanzar de nuevo decís «esta va ser la tercera cara de una serie de cinco caras seguidas». Después editáis el vídeo, de modo que incluya únicamente las

cinco caras seguidas. ¡Nadie se va a enterar! ¡Si verdaderamente queréis impresionar a la gente, podéis intentar con diez caras seguidas! (Hay más o menos un 38% de probabilidades de que eso suceda en una serie de 1.000 lanzamientos. O considerándolo desde otro punto de vista, si le pedimos a cien personas en una habitación que lancen una moneda cinco veces, hay un 96% de probabilidades de que uno de ellos saque cinco caras seguidas)[154].

El tipo de relación que un miembro de la alta sociedad de setenta y cinco años tiene con la policía de Nueva York probablemente es bastante diferente de la que tiene un chaval de color de dieciséis; sus experiencias están focalizadas selectivamente por lo que ven. Es posible que el joven de dieciséis años nos diga que lo paran a menudo sin causa aparente, en razón de su raza y que lo tratan como si fuera un criminal. El señor de setenta y cinco años puede que no comprenda cómo es eso posible. «La policía *siempre* me ha tratado *amablemente*».

Paul McCartney y Dick Clark compraron todo el metraje de televisión en el que aparecían durante los años sesenta, con la clara intención de controlar cómo se contaban sus respectivas historias[155]. Si sois unos académicos que hacen una investigación o unos documentalistas buscando metraje de archivo, vais a estar limitados a lo que ellos elijan proporcionaros. Cuando los datos o las pruebas apoyan una afirmación, preguntaos si lo que se os está mostrando es representativo del cuadro completo.

Pequeñas muestras selectivas

Las muestras pequeñas habitualmente no son representativas.

Suponed que sois los responsables de la campaña de marketing de un nuevo coche híbrido. Queréis afirmar su eficiencia energética. Enviáis a un conductor que utilice el vehículo y consigue hacer 128 kilómetros con 3,7 litros. ¡Es un resultado excelente, lo habéis conseguido! Pero puede que

hayáis tenido suerte. Vuestro competidor hace una prueba mayor, envía cinco conductores en cinco vehículos y obtiene una cifra que se aproxima más a los 96 kilómetros por 3,7 litros. ¿Quién está en lo cierto? ¡Ambos! Suponed que vuestro competidor muestra los resultados del siguiente modo:

Prueba 1: 92 km / 3,7 l.
Prueba 2: 60 km / 3,7 l.
Prueba 3: 111 km / 3,7 l.
Prueba 4: 87 km / 3,7 l.
Prueba 5: 128 km / 3,7 l.

Las condiciones de la carretera, la temperatura ambiente y los estilos de conducción generan una gran variabilidad. Si tuvisteis suerte (y vuestro competidor no), vuestro conductor puede haber obtenido un resultado extremo que podéis ofrecer con regodeo. (Si queréis hacer una selección interesada, podéis ignorar las pruebas 1-4). Pero si el investigador desea averiguar la verdad, necesita una muestra más amplia. Un laboratorio independiente que pusiera a prueba cincuenta trayectos diferentes podría averiguar que el promedio es completamente diferente. En general, es más fácil que se produzcan anomalías cuando las muestras son pequeñas. *Las muestras grandes reflejan con más rigor el mundo real*[156]. Los estadísticos lo denominan *ley de los grandes números*.

Si consideramos los nacimientos durante un mes en un pequeño hospital rural y observamos que el 70% de los bebés son niños, comparado con el 51% de un hospital urbano grande, podríamos creer que ocurre algo extraño en el hospital rural. Puede ser, pero no hay suficientes pruebas para estar seguro. Una vez más, la muestra pequeña en acción. El hospital grande puede que haya informado de que cincuenta y un nacimientos entre cien son niños, y el pequeño puede haber informado que son siete entre diez. Igual que pasaba con los lanzamientos de la moneda que comentamos antes, el promedio estadístico 50/50 es más fácil de obtener cuando las muestras son grandes.

¿Qué tamaño es suficiente? Eso es algo que ha de decidir un estadístico profesional, pero hay algunas reglas prácticas que podemos utilizar cuando intentamos comprender lo que leemos. Para sondeos de la población (intención de voto, preferencia por una pasta de dientes y cosas así), en la red se pueden encontrar calculadoras del tamaño de la muestra. Para determinar la incidencia local de algún suceso (como cuántos nacimientos son niños, cuántas veces tiene hambre al día el promedio de las personas), es necesario conocer algo sobre la tasa de base (o la tasa de incidencia) de lo que estamos buscando. Si un investigador deseara conocer cuántos casos de albinismo tienen lugar en una comunidad y analizase los primeros 1.000 nacimientos y no encontrara ninguno, sería estúpido concluir que el albinismo se da en uno de cada 17.000 nacimientos. Mil nacimientos es una muestra demasiado pequeña, «pequeña» en relación a la rareza de lo que estamos buscando. Por otra parte, si el estudio tratase de la incidencia de nacimientos prematuros[157], 1.000 debería resultar más que suficiente, ya que tienen lugar en uno de cada nueve nacimientos.

Alfabetización estadística

Imaginemos un juego callejero en el que un sombrero o un cesto contienen tres tarjetas, cada una con dos caras[158]: una de las tarjetas es roja por las dos caras, otra blanca por las dos caras y la otra roja por un lado y blanca por el otro. El timador saca una tarjeta del sombrero y os muestra una de las caras que es roja. Os apuesta 5 dólares a que la otra cara también es roja. Desea que creáis que la probabilidad es de un 50%, así que os tienta a apostar contra él, es decir, que hay la misma probabilidad de que la otra cara sea blanca. Es posible que razonéis del siguiente modo:

Me está enseñando una cara roja, así que o ha sacado la tarjeta roja-roja o la tarjeta roja-blanca. Eso significa que la otra cara es roja o blanca con la misma probabilidad. Puedo apostar, ya que si no gano en este envite, ganaré poco más tarde.

Dejando de lado la falacia del apostador —muchas personas han perdido dinero doblando la apuesta a la ruleta para acabar aprendiendo que el azar no es un proceso autocorrector—, el timador juega (¿apuesta?) a que vais a realizar esa estimación equivocada de la probabilidad y suele hablar rápido, para que tengáis que dividir vuestra atención. Es útil que lo veamos con una ilustración.

Estas son las tres tarjetas:

Rojo	Rojo	Blanco
Blanco	Rojo	Blanco

Si os está mostrando una cara roja, podría ser *cualquiera* de las tres caras rojas que hay. En dos de esos casos, la otra cara es roja y solo en un caso la otra cara es blanca. De modo que hay dos probabilidades sobre tres de que si os muestra la otra cara, esta sea roja y no una entre dos. Esto se debe a que la mayoría de nosotros no logramos tener en cuenta el hecho de que en el caso de la tarjeta de dos caras rojas, podría estarnos mostrando *cualquiera* de esas dos caras. Si lo anterior os hace sentir inquietos, no os atribuléis, el filósofo y matemático Gottfried Wilhelm Leibniz[159] y muchos autores de libros de texto más recientes cometían ese tipo de error. Cuando valoramos afirmaciones basadas en probabilidades, hay que comprender el modelo subyacente. Eso puede tener sus dificultades, pero si somos conscientes de que la probabilidad puede ser engañosa y reconocemos las limitaciones que la mayoría de nosotros tenemos para valorarla, será más difícil que nos engañen. Pero ¿qué pasaría si todos los que nos rodean creyeran algo que está errado? ¿El traje nuevo del emperador, tal vez?

---

145 W. Durant y A. Durant (1968), *The Lessons of History.* Nueva York, Simon & Schuster.

146 Federal Bureau of Investigation (20 de abril de 2015), «FBI testimony on microscopic hair analysis contained errors in at least 90 percent of cases in ongoing review [Press release]», https://www.fbi.gov/news/pressrel/press-releases/fbi-testimony-on-microscopic-hair-analysis-contained-errors-in-at-least-90-percent-of-cases-in-ongoing-review.

147 C. G. G. Aitken y F. Taroni (2004), *Statistics and the Evaluation of Evidence for Forensic Scientists,* 2.ª ed., Chicester, UK, John Wiley & Sons, p. 95, citando a R. D. Friedman (1996), «Assessing Evidence», *Michigan Law Review,* 94(6), 1810-1838.

148 R v. Dennis John Adams (1996), 2 Cr App R, 467; y C. Aitken (2003), «Statistical techniques and their role in evidence interpretation», en J. Payne-James, A. Busuttil y W. Smock (eds.), *Forensic Medicine: Clinical and Pathological Aspects,* Cambridge, UK, Cambridge University Press.

149 R. Blumenthal (3 de noviembre de 2015), «Built by the ancients, seen from space», *New York Times,* p. D2.

150 Agradezco a Stephen Kosslyn haber compartido conmigo una versión de este ejemplo.

151 W. Grimes (13 de noviembre de 2015), «Jack Yufe, a Jew whose twin was a nazi, dies at 82», *New York Times,* p. B8.

152 Buena parte de esto está tomado literalmente de Grimes (2015), *op. cit.*

153 Dr. Jeffrey Mogil, comunicación personal.

154 La fórmula es $1 - (1 - \tfrac{1}{2}^{5})^{100}$.

155 Agradezco a Ron Mann esta observación.

156 Obsérvese que en una muestra grande resulta más probable que encontremos una observación anómala (valor atípico) que en una muestra pequeña, pero cuando consideramos la *media,* la media de una muestra grande es mucho más probable que refleje la situación real (ya que hay tantas observaciones más que anulan la anomalía).

157 W. Krämer y G. Gigerenzer (2005), «How to confuse with statistics or: the use and misuse of conditional probabilities», *Statistical Science,* 20(3), 223-230. Véase también Centers for Disease Control and Prevention, «Preterm birth», http://www.cdc.gov/reproductivehealth/maternalinfanthealth/pretermbirth.htm.

158 W. Krämer y G. Gigerenzer (2005). Técnicamente, según señalan, esta es una enumeración incorrecta de sucesos simples en un experimento laplaciano de la subpoblación compuesta por las posibilidades restantes.

159 *Ibid*.

CONTRACONOCIMIENTO

El contraconocimiento, un término acuñado por el periodista británico Damian Thompson[160], es la información errónea dispuesta para que parezca un hecho y que es creída por una cantidad significativa de personas. Un presidente de Estados Unidos recién elegido, habiendo ganado el voto del colegio electoral[161], afirmó que también había ganado el voto popular, cuando se disponía de abundante y sólida evidencia de que no era así. El contraconocimiento se fue repitiendo. Poco después una encuesta reveló que el 52% de quienes apoyaban al presidente, decenas de millones de personas, creían tal mentira. No es solo en la política donde se propaga el contraconocimiento. Hay ejemplos en la ciencia, en los temas de actualidad, en el cotilleo sobre los famosos y en la pseudohistoria. En ocasiones se divulgan inmensas conspiraciones, como cuando se afirma que el Holocausto, los alunizajes y los ataques contra Estados Unidos del 11 de septiembre de 2001 nunca sucedieron. A veces es la absurda afirmación de que una pizzería sirve como tapadera de un negocio de sexo infantil dirigido por la anterior secretaria de Estado.

La intrigante idea de imaginar «¿y si fuera verdad?» contribuye a la difusión de las historias falsas. Insisto, los humanos somos una especie dada a la narrativa, nos fascina un buen cuento. El contraconocimiento nos atrae inicialmente con una pátina de conocimiento y autoridad, pero un examen más minucioso muestra que no existe un fundamento real, quienes generan el contraconocimiento esperan que quedemos suficientemente impresionados (o intimidados) por la presencia de afirmaciones y cifras descarnadas y que las vamos a aceptar ciegamente.

Damian Thompson[162] describe cómo es posible que esas afirmaciones arraiguen, se nos metan bajo la piel y provoquen que dudemos de lo que sabemos... es decir, hasta que les

aplicamos un análisis racional. Thompson recuerda la ocasión en que un amigo, hablando sobre los ataques del 11 de septiembre en Estados Unidos, «atrajo nuestra atención con una observación de aroma verosímil: “Fijaos cómo las torres se desplomaron verticalmente, en lugar de caer de lado. El combustible aeronáutico no es capaz de generar el calor suficiente para derretir el acero. Únicamente una explosión controlada puede hacer algo así”».

La anatomía de este contraconocimiento es más o menos esta:

Las torres se desplomaron verticalmente: eso es verdad, hemos visto las grabaciones.

Si el ataque se hubiera producido como nos han contado, los edificios hubieran debido caer de lado: esa es una premisa no explícita, oculta. No sabemos si eso es verdad. El mero hecho de que nuestro interlocutor lo afirme no hace que eso sea verdad. Se trata de un aserto que requiere verificación.

El combustible aeronáutico no puede generar el calor suficiente para derretir el acero: tampoco sabemos si eso es verdad y se ignora el hecho de que es posible que también hubiera otros materiales inflamables, productos de limpieza, pintura, productos químicos industriales, en el edificio, de modo que cuando el fuego comenzó se sumaron al incendio.

Si no sois ingenieros de estructuras profesionales, es posible que creáis que esas premisas son verosímiles. Pero una breve indagación revela que los ingenieros de estructuras profesionales no encuentran nada misterioso en el desplome de las torres. No os dejéis influenciar por las opiniones de pseudoexpertos en temas colaterales que no son especialistas en cómo se desploman los edificios.

Es importante darse cuenta de que en los sucesos complejos no todo es explicable, ya que no es posible observarlo todo o informar de todo. En el asesinato del presidente John F. Kennedy la grabación Zapruder es la única prueba fotográfica de la secuencia del suceso y está incompleta. Se grabó con una cámara de aficionado[163] a 18,3 imágenes por segundo y con baja resolución. Hay muchas preguntas sin responder sobre el asesinato y algunas indicaciones de que las pruebas se

manipularon, muchos testigos presenciales nunca fueron interrogados y muchas muertes por explicar de personas que afirmaban saber o creían saber lo que realmente ocurrió. Es posible que haya habido una conspiración, pero el mero hecho de que haya preguntas sin responder o inconsistencias no prueba que la hubiera. Un dolor de cabeza con la vista borrosa no es la prueba de que tenemos un raro tumor cerebral, lo más probable es que se trate de algo bastante menos grave.

Los científicos y otros pensadores racionales diferencian entre aquello que sabemos que es verdad de modo casi seguro, como la fotosíntesis o que la Tierra gira alrededor del Sol y cosas que son *probablemente* ciertas, como que los ataques del 11 de septiembre fueron el resultado del secuestro de varios aviones y no un complot del gobierno de Estados Unidos. Existen diferentes cantidades de pruebas y diferentes tipos de pruebas sobre cada uno de estos asuntos. Que una teoría o una explicación presente lagunas no la desacredita. Un *puñado* de anomalías sin explicar[164] no desacredita ni socava una teoría bien establecida que se funda en *miles* de pruebas. Sin embargo, esas anomalías forman el meollo típico del pensamiento conspiratorio, del revisionismo del Holocausto, del antievolucionismo y de las teoría de la conspiración del 11 de septiembre. La diferencia entre una teoría falsa y una teoría verdadera es la probabilidad[165]. Thompson designa algo como contraconocimiento cuando es contrario al conocimiento real y cuenta con cierto apoyo social.

Cuando los periodistas nos confunden

Los periodistas obtienen información sobre los sucesos importantes de dos modos diferentes. A menudo esos dos modos son incompatibles entre sí y dan lugar a historias que pueden engañar al público si los periodistas no son cuidadosos.

En el modo de *investigación científica,* los periodistas colaboran con los científicos, informan sobre descubrimientos científicos y procuran comunicarlos en un lenguaje que el público pueda entender, algo que la mayoría de los científicos no hacen

bien. El periodista lee sobre una investigación en una revista con revisión de expertos o en una nota de prensa. Cuando un estudio alcanza el punto de ser revisado por expertos habitualmente es analizado por entre tres y cinco científicos especializados que dan el visto bueno a su exactitud y a sus conclusiones. Habitualmente no es tarea del periodista establecer la solidez de las pruebas científicas que apoyan las hipótesis, las hipótesis auxiliares y las conclusiones, eso ya lo han hecho los científicos que revisan el artículo.

En ese punto el proceso se divide entre dos tipos de periodistas. Los periodistas de investigación serios, como los del *Washington Post* o el *Wall Street Journal,* habitualmente toman contacto con algunos científicos *no relacionados* con la investigación en cuestión para obtener sus opiniones. Lo que se busca son opiniones contrarias al artículo publicado. Pero la mayoría de los periodistas consideran que su trabajo se reduce a informar sobre el estudio tal como se ha publicado, trasladándolo a un lenguaje más sencillo.

En el modo *noticias de última hora* los periodistas procuran descubrir algo que está sucediendo en el mundo reuniendo información de distintas fuentes, testigos de los hechos. Puede ser alguien que presenció un atraco en Detroit o un bombardeo en Gaza o una concentración de tropas en Crimea. El periodista puede disponer de un único testigo o intentar corroborar con un segundo o un tercero. Parte del trabajo del reportero en esos casos consiste en asegurarse de la veracidad y fiabilidad del testigo. Preguntas como «¿lo vio usted mismo?» o «¿dónde estaba usted cuando esto sucedió?» le ayudan a hacerlo. Os sorprendería saber lo frecuente que es que la respuesta sea no, o cuán a menudo las personas mienten y tan solo mediante una verificación cuidadosa logran los reporteros que las inconsistencias salgan a la luz.

Por lo tanto, en el Modo Uno los periodistas informan sobre descubrimientos científicos que probablemente se basan en miles de observaciones y en una gran cantidad de datos. En el Modo Dos los periodistas informan sobre sucesos que frecuentemente se fundamentan en lo que describen unos pocos testigos.

Dado que los periodistas han de trabajar empleando ambos modos, a veces los confunden. A veces se olvidan de que el plural de anécdota no es datos; es decir, que un puñado de historias o de observaciones informales no son ciencia. A este hecho se suma nuestra expectativa de que, además de aprender, los periódicos deberían entretenernos, contarnos historias. La mayoría de las buenas historias nos muestran una cadena de acciones que pueden relacionarse en términos de causa y efecto. Las hipotecas de riesgo fueron reembaladas como productos de inversión calificados como AAA, lo que llevó a reventar la burbuja inmobiliaria en 2007. Los reguladores ignoraron la montaña de escombros sobre la ciudad china de Shenzhen y en 2015 se desplomó y formó una avalancha que derribó treinta y tres edificios. No se trata de experimentos científicos, sino de sucesos que intentamos comprender para elaborar una narración al respecto. La carga de la prueba es distinta para los artículos periodísticos y para los científicos, pero si carecemos de una explicación, siquiera sea de una tentativa, no tenemos ninguna historia. Y, al igual que sucede con los periódicos, las revistas y los libros, las personas necesitan historias.

Esta es la razón por la que los rumores, el contraconocimiento y los pseudohechos pueden propagarse tan fácilmente en los medios de comunicación, como cuando Geraldo Rivera contribuyó a crear un pánico nacional sobre el control satánico de Estados Unidos en 1987[166]. Hemos sufrido varios pánicos semejantes en los medios de comunicación sobre la abducción por extraterrestres y los recuerdos reprimidos. Tal como señala Damian Thompson, «para el editor en apuros un testimonio angustioso siempre es mejor que una estadística árida y probablemente no concluyente»[167]. En la mayoría de las narraciones periodísticas y de los descubrimientos científicos no existe la certeza absoluta. Pero como humanos buscamos la certeza. Los demagogos, los dictadores, las sectas e incluso algunas religiones nos la ofrecen —una falsa certeza—, y muchas personas la encuentran irresistible.

Percepción del riesgo

Tendemos a suponer que el espacio que los medios de comunicación conceden a la información delictiva es un indicador de la tasa de criminalidad. O que la cobertura que conceden los periódicos a diversas causas de muerte guarda una relación con los riesgos. Sin embargo, esas suposiciones son imprudentes. Cada año mueren aproximadamente cinco veces más personas de cáncer de estómago[168] que por ahogamiento accidental[169]. Pero echad un vistazo simplemente a un periódico, el *Sacramento Bee* informó de tres ahogamientos accidentales en 2014, pero de ningún cáncer de estómago. Si nos basáramos en la cobertura que se concede a las noticias, pensaríamos que los ahogamientos accidentales son mucho más frecuentes que las muertes por cáncer de estómago. El psicólogo cognitivo Paul Slovic ha demostrado que la gente concede una relevancia exagerada a los riesgos relativos que reciben atención en los medios de comunicación. Parte de lo que provoca que algo reciba atención en los medios de comunicación es si constituye o no una buena historia. Una muerte por ahogamiento es más dramática, más repentina y tal vez más fácil de evitar que una muerte por cáncer de estómago, todos ellos son elementos con los que se puede construir un buen cuento, aunque sea trágico. Por lo tanto, se informa más sobre muertes por ahogamiento y eso nos lleva a creer erróneamente que son más frecuentes. No comprender los riesgos puede conducirnos a ignorar o a descartar datos que podríamos utilizar para protegernos (véase la figura de la página siguiente).

Empleando este principio de la falta de comprensión de los riesgos, un estadístico aficionado o meramente carente de escrúpulos y con acceso algún medio de comunicación puede embaucarnos, haciéndonos creer cosas que no existen.



«Te lo tomas muy a la ligera teniendo en cuenta todo lo que se ha hablado de los tsunamis en la prensa».

Un titular de primera página del *Times* (británico)[170] anunciaba, en 2015, que los británicos que contraerían un cáncer durante sus vidas pasaría de un 33 a un 50%. Esto supondría que dos tercios de quienes son niños hoy en día se verían afectados, lo que implica el riesgo de que la Seguridad Social se viera inundada por la cantidad de pacientes de cáncer. ¿Qué nos lleva a pensar esa noticia? ¿Quizá se está desatando una epidemia de cáncer? Puede que haya algo que sospechar en el estilo de vida moderno, la comida basura insalubre, la radiación que emiten los teléfonos móviles, los productos de limpieza carcinógenos y la radiación que atraviesa el agujero de ozono. (Aquí entra en juego el pensamiento creativo que comenté en la introducción). No cabe duda de que el titular podría utilizarse para promover los planes de diversos grupos interesados por lucrarse, empresas de comida

saludable, fabricantes de protectores solares, practicantes de la medicina holística e instructores de yoga.

Pero antes de entrar en pánico, daos cuenta de que esa cifra incluye todos los tipos de cáncer, entre los que se encuentran los de desarrollo lento, como el cáncer de próstata, los melanomas que se pueden eliminar fácilmente, etc. No significa que todos los que contraigan un cáncer van a morir. La Asociación de Investigación del Cáncer británica[171] nos informa de que el porcentaje de personas que logran vencer al cáncer se ha doblado desde los años setenta, gracias al diagnóstico temprano y a la mejora de los tratamientos.

Lo que el titular no cuenta es que gracias a los avances de la medicina, las personas viven más tiempo. La enfermedad cardíaca se está controlando mejor que nunca[172] y las muertes por enfermedades respiratorias han disminuido radicalmente en los últimos veinticinco años. La principal razón por la que tanta gente muere de cáncer es que, para empezar, no se están muriendo por otras causas. Todos hemos de morir de *algo*. Esta idea estaba incluida en el artículo del *Times,* si uno llegaba a leer hasta ese punto (cosa que muchos de nosotros no hacemos, nos detenemos en el titular, nos asustamos y nos preocupamos). Parte de lo que refleja la estadística a la que alude el titular es que el cáncer es una enfermedad de ancianos y muchos de nosotros viviremos lo suficiente para contraerlo. Pero eso no necesariamente debe llevarnos al pánico. Lo anterior sería equivalente a decir «la mitad de los coches de Argentina tendrán un fallo completo del motor durante la vida útil del vehículo». Claro, sin duda el coche tendrá que retirarse del servicio por algún motivo. Podría ser un eje roto, una colisión grave, una transmisión que no funciona o un fallo del motor, pero algo tiene que ser la causa.

Persuasión por asociación

Si pretendemos inundar a la gente con contraconocimiento, una técnica eficaz consiste en coger un puñado de hechos reales y

verificables y añadirles uno o dos que sean falsos. Contaremos con los hechos reales para darles un aire de verdad y los intrépidos exploradores de Internet que procuren verificarlos tendrán éxito. Así que nos limitamos a añadir una o dos mentiras para llevarnos el gato al agua y muchos desgraciados nos creerán. Conseguimos persuadir mediante la asociación de hechos falsos o contraconocimiento con hechos reales y conocimiento verídico.

Veamos el siguiente argumento:

1. El agua está formada por hidrógeno y oxígeno.
2. El símbolo molecular del agua es H_2O.
3. Nuestros cuerpos están constituidos por más de un 60% de agua.
4. La sangre humana es agua en un 92%.
5. El cerebro es un 75% de agua.
6. Muchos lugares del planeta tienen agua contaminada.
7. Menos del 1% del agua accesible en el mundo es potable.
8. Únicamente podemos estar seguros de la calidad del agua si compramos agua embotellada.
9. Los principales investigadores sanitarios recomiendan beber agua embotellada y la mayoría de ellos beben agua embotellada.

Los enunciados del uno al siete son todos verdad. El enunciado ocho no es una conclusión que se siga lógicamente y en cuanto al nueve, bueno ... ¿quiénes son los principales investigadores sanitarios? A lo mejor en una fiesta, o en un restaurante, o en un avión, cuando sirven agua y no hay otra alternativa, se la beben. ¿O significa que evitan escrupulosamente cualquier otra forma de agua? Hay una enorme diferencia entre estas dos posibilidades.

El hecho es que en el mejor de los casos el agua embotellada no es más segura que el agua del grifo en la mayoría de los países desarrollados y, en algunos casos, es menos segura a causa de normativas menos estrictas. Lo anterior se fundamenta en informes realizados por diversas fuentes reputadas[173], entre las que se incluyen el Natural Resources Defense Council, la Clínica Mayo, *Consumer Reports*[174] y bastantes trabajos publicados en revistas con revisión por expertos.

No cabe duda de que hay excepciones. En la ciudad de Nueva York, en Montreal, en Flint, Michigan y en muchas otras ciudades más antiguas[175], el agua se conduce por tuberías de plomo y este puede contaminar el agua del grifo y provocar envenenamientos. La existencia de problemas periódicos en las plantas de tratamiento ha provocado que algunos ayuntamientos no recomienden beber el agua del grifo. Cuando se viaja a países del Tercer Mundo en los que la normativa y los niveles sanitarios son menos estrictos, la opción más segura puede ser beber agua embotellada. Pero el nivel del agua del grifo en los países industrializados cumple requisitos más estrictos que cualquier industria, ahorrad vuestro dinero y evitad las botellas de plástico. En fin, que el argumento de los defensores pseudocientíficos de la salud, tal como se ha expuesto más arriba, no se sostiene.

---

160 D. Thompson (2008), *Counterknowledge: How We Surrendered to Conspiracy Theories, Quack Medicine, Bogus Science, and Fake History,* Nueva York, W. W. Norton, p. 1.

161 *N. del T.:* El sistema para las elecciones presidenciales en Estados Unidos es indirecto, los ciudadanos eligen un colegio electoral, compuesto por 538 electores, para ser elegido presidente son necesarios los votos de 270 de éstos. A cada estado le corresponde una cuota de tales electores. Es posible obtener mayoría en el colegio electoral, aunque no se alcance en el voto popular directo.

162 D. Thompson (2008), *op. cit.*

163 R. B. Trask (1996), *Photographic Memory: The Kennedy Assassination, November 22, 1963,* Dallas, Sixth Floor Museum, p. 5.

164 Agradezco a Michael Shermer esta observación.

165 Cita literal de D. Thompson (2008), *op. cit.*

166 *N. del T.:* Geraldo Rivera es un conocido presentador de televisión en Estados Unidos, uno de cuyos magazines informativos contribuyó a desatar un pánico satánico en ese país, al afirmar que centenares de miles de ciudadanos estarían organizados en una compleja red de organizaciones satánicas.

167 D. Thompson (2008), *op. cit.,* p. 17. Las frases previas son de las pp. 16-17.

168 National Cancer Institute, «SEER stat fact sheets: stomach cancer», http://seer.cancer.gov/statfacts/html/stomach.html.

169 Centers for Disease Control and Prevention, «Unintentional drowning: get the facts», http://www.cdc.gov/HomeandRecreationalSafety/WaterSafety/waterinjuries-factsheet.html.

170 C. Smyth (4 de febrero de 2015), «Half of all Britons will get cancer during their lifetime», *Times,* www.thetimes.co.uk/tto/health/news/article434681.ece.

171 S. Boseley (3 de febrero de 2015), «Half of people in Britain born after 1960 will get cancer, study shows», *Guardian.*

172 C. Griffths y A. Brock (2003), «Twentieth century mortality trends in England and Wales», *Health Statistics Quarterly,* 18(2), 5-17.

173 http://www.nrdc.org/water/drinking/qbw.asp; http://www.mayoclinic.org/healthy-lifestyle/nutrition-and-healthy-eating/expert-answers/tap-vs-bottled-water/faq-20058017; http://www.consumerreports.org/cro/news/2009/07/is-tap-water-safer-than-bottled/index.htm; http://news.nationalgeographic.com/news/2010/03/100310/why-tap-water-is-better/; http://abcnews.go.com/Business/study-bottled-water-safer-tap-water/story?id=87558; http://www.telegraph.co.uk/news/health/news/9775158/Bottled-water-not-as-safe-as-tap-variety.html.

174 *N. del T.:* El NRDC es una organización sin ánimo de lucro que agrupa a millones de asociados y centenares de científicos y tiene por misión asegurar el acceso de la población a los recursos naturales. La Clínica Mayo es una organización sin ánimo de lucro de gran prestigio en el ámbito de la medicina. *Consumer Reports* es una famosa revista de la Unión de Consumidores en Estados Unidos.

175 N. Stockton (29 de enero de 2016), «Here's how hard it will be to unpoison Flint's water», *Wired,* http://www.wired.com/2016/01/heres-how-hard-it-will-be-to-unpoison-flints-water/.

PARTE III

EVALUAR EL MUNDO

La naturaleza nos permite únicamente calcular probabilidades. Sin embargo, la ciencia no ha colapsado[176].

Richard P. Feynman

---

176 R. P. Feynman (1985), *QED: The Strange Theory of Light and Matter,* Princeton, NJ, Princeton University Press.

CÓMO FUNCIONA LA CIENCIA

El desarrollo del pensamiento crítico a lo largo de muchos siglos ha dado lugar a un cambio de paradigma en el pensamiento humano y en la historia: la revolución científica. Sin su desarrollo y su práctica en ciudades como Florencia, Bolonia, Gotinga, París, Londres y Edimburgo, por nombrar nada más que un puñado de grandes centros del conocimiento, la ciencia podría no haber conformado nuestra cultura, nuestra industria y nuestras mayores ambiciones. Por supuesto, la ciencia no es infalible, pero el pensamiento científico representa la base de gran parte de lo que hacemos y de cómo tratamos de decidir lo que existe y lo que no. Esto hace que merezca la pena echar un buen vistazo detrás del telón para tratar de entender qué hace y cómo lo hace. Eso incluye ver cómo nuestros imperfectos cerebros humanos, incluso los de los pensadores más rigurosos, pueden engañarse a sí mismos.

Lamentablemente, también debemos reconocer que algunos investigadores se inventan los datos. En los casos más extremos, informan de datos que jamás han sido recogidos de experimentos que nunca se han llevado a cabo. Se salen con la suya porque el fraude es relativamente raro entre investigadores y, por tanto, los evaluadores expertos están desprevenidos. En otros casos, un investigador cambia algunos datos para hacer que el conjunto refleje de manera más precisa su hipótesis favorita. En los casos menos extremos, el investigador omite algunos datos porque no concuerdan con la hipótesis, o elige solamente casos que sabe que apoyarán dicha hipótesis. Un caso de fraude ocurrió en 2015 cuando se descubrió que Dong-Pyou Han[177, 178], un antiguo investigador biomédico de la universidad del Estado de Iowa en Ames, había falsificado e inventado información acerca de una potencial vacuna contra el VIH. Como resultado inesperado no solo perdió su trabajo en la universidad, sino que además fue

sentenciado a casi 5 años de prisión. Toda la controversia sobre si la vacuna del sarampión, paperas y rubeola (SPR) causaba autismo fue propagada por Andrew Wakefield en un artículo con datos falsificados que ahora ha sido denunciado y retirado —y aun así millones de personas continúan creyendo en esta relación. En algunos casos, un investigador manipula o elimina datos de acuerdo con principios establecidos, pero omite indicar que lo ha hecho, lo cual implica que la interpretación y la replicación sean más difíciles (y está en el límite de la mala praxis científica).

La búsqueda de pruebas o de certezas rige la ciencia, pero también rige nuestro sentido de la justicia y todos nuestros sistemas judiciales. La práctica científica nos ha enseñado el camino correcto por el que proceder en esta búsqueda.

Hay dos mitos extendidos sobre cómo se hace la ciencia: el primero es que la ciencia es clara y ordenada, que los científicos nunca discrepan en nada; el segundo es que un solo experimento nos dice todo lo que necesitamos saber sobre un fenómeno, que la ciencia avanza a pasos agigantados después de la publicación de cada experimento. La ciencia real está repleta de controversia, duda y debate acerca de lo que verdaderamente conocemos. El auténtico conocimiento científico se establece gradualmente, a través de muchas réplicas y descubrimientos convergentes. El conocimiento científico surge de acumular grandes cantidades de datos a partir de muchos experimentos realizados por múltiples laboratorios. Cualquier experimento en concreto es solo un ladrillo en un gran muro. Únicamente cuando una masa crítica de experimentos ha sido completada estamos en posición de observar el muro completo de los datos y de sacar cualquier conclusión firme.

La divisa vigente no es el experimento único, sino el metaanálisis. Previo al consenso científico acerca de algo, habitualmente ha habido un metaanálisis que reúne las distintas pruebas a favor o en contra de una hipótesis.

La idea del metaanálisis frente al experimento único debería recordaros los problemas del foco selectivo y el problema de las muestras pequeñas de la segunda parte. Un único experimento,

incluso con muchos participantes u observaciones, podría ser meramente una anomalía —esos 34 km por litro que tuvisteis la fortuna de alcanzar cuando probasteis el coche. Una docena de experimentos llevados a cabo en distintos momentos y lugares os darán una idea más clara de hasta qué punto es sólido el fenómeno. La próxima vez que leáis sobre una nueva crema facial que os hará rejuvenecer veinte años, o sobre un nuevo remedio de hierbas para el resfriado común, una de las cosas que habría que preguntar es si un metaanálisis está respaldando la afirmación o si es un único estudio.

Deducción e inducción

El progreso científico depende de dos tipos de razonamiento: mediante la deducción razonamos de lo general a lo específico y si seguimos las reglas de la lógica, podemos estar seguros de nuestra conclusión; mediante la inducción tomamos una serie de observaciones o hechos y tratamos de obtener un principio general que los explique. Es un razonamiento de lo específico a lo general. La conclusión del razonamiento inductivo no es segura, está basada en nuestras observaciones y entendimiento del mundo, e implica ir un paso más allá de lo que nos dicen los datos.

La probabilidad, como indicamos en la primera parte, es deductiva; trabajamos desde información general (como «esto es una moneda sin trucar») a una predicción específica (la probabilidad de sacar cara tres veces seguidas). La estadística es inductiva; trabajamos desde un conjunto particular de observaciones (como sacar cara tres veces seguidas) a un principio general (sobre si la moneda está trucada o no). Otro ejemplo podría ser utilizar la probabilidad (deducción) para establecer la expectativa de que un medicamento concreto para el dolor de cabeza nos alivie. Si el dolor de cabeza no desaparece podríamos usar la estadística (inducción) para estimar la probabilidad de que la pastilla venga de un mal lote.

La inducción y la deducción no solo se aplican a objetos numéricos, como la probabilidad o la estadística. Aquí tenéis un ejemplo de lógica deductiva en palabras. Si la premisa (el primer enunciado) es cierta, la conclusión también ha de serlo:

> Gabriel García Márquez es humano.
> Todos los humanos son mortales.
> Por tanto (esto es una conclusión deductiva), Gabriel García Márquez es mortal.

1. Algunos automóviles son Ford.
2. Todos los Ford son automóviles.
3. El tío que hizo de Han Solo es un automóvil.



El tipo de argumento deductivo sobre Márquez se denomina silogismo. En un silogismo la *forma* del argumento garantiza la conclusión que le sigue. Se puede construir un silogismo con una premisa que sepamos (o creamos) que es falsa, pero eso no invalida el silogismo, es decir, la lógica del asunto se mantiene.

> La Luna está hecha de queso verde.
> El queso verde cuesta 22,99 dólares por libra.
> Por tanto, la Luna cuesta 22,99 dólares por libra.

Claramente la Luna *no* está hecha de queso verde, pero SI lo estuviera, la deducción lógica sería válida. Si os hace sentir

mejor, podemos reescribir el silogismo de tal manera que resulte explícito:

SI la Luna está hecha de queso verde
Y SI el queso verde cuesta 22,99 dólares por libra
ENTONCES la Luna cuesta 22,99 dólares por libra.

Hay varios tipos distintos de argumento deductivo y normalmente se enseñan en clases de filosofía o de matemáticas sobre lógica formal. Otra forma común implica a los condicionales; este se llama *modus ponens*. Es fácil de recordar cómo se llama usando este ejemplo (usando Poe, como en *ponens*):

Si Edgar Allan Poe fue a la fiesta, llevó puesta una capa negra.
Edgar Allan Poe fue a la fiesta.
Por tanto, llevó puesta una capa negra.

Puede costar un tiempo dominar la lógica formal porque, al igual que sucede con muchas otras formas de razonar, nuestras intuiciones nos fallan. ¿Suena lo siguiente como una conclusión válida? ¿O no válida?

Si Edgar Allan Poe fue a la fiesta, llevó puesta una capa negra.
Edgar Allan Poe llevó puesta una capa negra.
Por tanto, fue a la fiesta.

Si bien *podría* ser cierto que Poe fuese a la fiesta, no es *necesariamente* cierto. Podría haber llevado la capa por otra razón (quizá hacía frío, igual era *Halloween,* quizá estaba actuando en una obra que requería una capa y quería meterse en el personaje). Sacar la conclusión arriba expuesta representa un error de razonamiento llamado *falacia de afirmación del consecuente* o *error de conversión*.

Si os resulta difícil recordar cómo se llama, considerad este ejemplo:

Si Chuck Taylor lleva unas Converse, entonces sus pies están calzados.
Los pies de Chuck Taylor están calzados.
Por tanto, lleva puestas unas Converse.

Obviamente este razonamiento no se sostiene, porque llevar unas Converse no es la única manera de tener los pies calzados, podríais haber llevado puestos muchos tipos de marcas de zapatos diferentes, o tener bolsas de basura en los pies atadas alrededor de los tobillos.

Sin embargo, *podemos* decir con seguridad que si los pies de Chuck Taylor no están calzados, seguro que no lleva unas Converse. A esto se le llama el *contrapositivo* del primer enunciado.

Los enunciados lógicos no funcionan como los símbolos de restar en las ecuaciones, no se puede simplemente negar un lado y hacer que esto niegue automáticamente el otro. Es necesario memorizar estas reglas. Resulta algo más fácil de hacer utilizando una notación cuasimatemática. Los enunciados anteriores se pueden representar de la siguiente manera: A sería cualquier premisa, como «si Chuck Taylor lleva unas Converse» o «si la Luna está hecha de queso verde» o «si los Mets se alzan victoriosos este año». B sería la consecuencia como en «entonces los pies de Chuck están calzados» o «entonces la Luna debería ser verde en el cielo nocturno» o «me comeré mi sombrero».

Usando esta notación generalizada, decimos *Si A* como una abreviatura de «si A es cierto». Decimos *B* o *No B* como abreviatura de «B es cierto» o «B no es cierto». Entonces...

Si A, entonces B
A
Entonces, B

En los libros de lógica podemos encontrar la palabra *entonces* reemplazada por una flecha (→) y podemos ver la palabra *no* reemplazada por este símbolo: ~. Podemos ver *por tanto* reemplazado por (∴) como en:

Si A → B
A
∴ B

No dejéis que esto os inquiete. Solo son personas tratando de ser sofisticadas.

Pues bien, hay cuatro posibilidades para enunciados como este: A puede ser cierto o no cierto, y B puede ser cierto o no cierto. Cada una de las posibilidades tiene un nombre especial.

1. Modus ponens. A esto también se le llama afirmación del antecedente.

   Si A → B

   A ∴ B



Ejemplo: Si esa mujer es mi hermana, entonces es más joven que yo.
Esa mujer es mi hermana.
Por tanto, es más joven que yo.

2. El contrapositivo.

   Si A → B

   ~ B ∴ ~ A



Ejemplo: Si esa mujer es mi hermana, entonces es más joven que yo.
Esa mujer no es más joven que yo.
Por tanto, no es mi hermana.

3. El contrario.

   Si A → B

   B ∴ A



Esto *no* es una deducción válida.

Ejemplo: Si esa mujer es mi hermana, entonces es más joven que yo.
Esa mujer es más joven que yo.

Por tanto, es mi hermana.

Esto no es válido porque hay muchas mujeres más jóvenes que yo que no son mi hermana.

4. El inverso.

   Si A → B

   ~A ∴ ~B



Esto *no* es una deducción válida.

Ejemplo: Si esa mujer es mi hermana, entonces es más joven que yo.

Esa mujer no es mi hermana.

Por tanto, no es más joven que yo.

Esto no es válido porque muchas mujeres que no son mi hermana aun así son más jóvenes que yo.

El razonamiento inductivo está basado en la existencia de pruebas que sugieran que la conclusión es cierta, pero no lo garantizan. Al contrario que en la deducción, esto lleva a conclusiones inciertas, pero probables (si se hace correctamente).

Un ejemplo de inducción es:

Todos los mamíferos conocidos hasta ahora tienen riñones.

Por tanto (este es el paso inductivo), si descubrimos un nuevo mamífero, probablemente tendrá riñones.

La ciencia progresa por una combinación de deducción e inducción. Sin inducción, no tendríamos ninguna hipótesis sobre el mundo. La usamos constantemente en el día a día.

Cada vez que he contratado a Patrick para hacer una reparación en la casa ha hecho una chapuza.

Por tanto, si contrato a Patrick para hacer la próxima reparación, también hará una chapuza.

Todos los pilotos de aerolínea que he conocido son organizados, concienzudos y meticulosos.

Lee es piloto de aerolínea. Él tiene esas cualidades y además es bueno en matemáticas.

Por tanto, todos los pilotos de aerolínea son buenos en matemáticas.

Por supuesto, este segundo ejemplo no tiene por qué cumplirse. Estamos haciendo una inferencia. Con lo que sabemos del mundo y los requisitos profesionales para ser piloto —trazar rutas, estimar la influencia de la velocidad del viento en el tiempo de llegada, etc.—, esto parece razonable. Pero consideremos:

Todos los pilotos de aerolínea que he conocido son organizados, concienzudos y meticulosos.

Lee es piloto de aerolínea. Él tiene esas cualidades y también le gusta la fotografía.

Por tanto, a todos los pilotos de aerolínea les gusta la fotografía.

Aquí nuestra inferencia es más cuestionable. Nuestro conocimiento del mundo real sugiere que la fotografía es una preferencia personal, y no se cumple necesariamente que a un piloto le guste más o menos que a un no piloto.

El gran detective de ficción Sherlock Holmes saca conclusiones mediante razonamientos perspicaces y aunque afirma estar usando la deducción, en realidad está usando una forma diferente de razonamiento llamado *abducción*. Casi todas las conclusiones de Holmes son suposiciones perspicaces basadas en hechos, pero no de manera que la conclusión sea irrefutable o inevitable. En el razonamiento abductivo, empezamos con un conjunto de observaciones y después generamos una teoría que las explica. De la infinidad de teorías diferentes que podrían explicar algo, buscamos la más probable.

Por ejemplo, Holmes concluye que un supuesto suicidio ha sido un asesinato[179]:

HOLMES: La herida está en el lado derecho de su cabeza. Van Coon era zurdo. Requeriría bastante contorsión.

DETECTIVE INSPECTOR DIMMOCK: ¿Zurdo?

HOLMES: Oh, me sorprende que no lo haya notado. Lo único que tiene que hacer es mirar alrededor del apartamento. Mesita del café en el lado izquierdo; asa de la taza de café apuntando hacia la izquierda. Enchufes: los izquierdos habitualmente usados... Papel y lápiz a la izquierda del teléfono porque lo cogía con la derecha y tomaba los recados con la izquierda... Hay un cuchillo en la tabla de cortar con mantequilla en el lado derecho del filo porque lo usaba con la izquierda. Es altamente improbable que un hombre zurdo se disparara en el lado derecho de su cabeza. Conclusión: alguien ha entrado por la fuerza y lo ha asesinado...

DIMMOCK: Pero el arma... por qué...

HOLMES: Estaba esperando al asesino. Había sido amenazado.

Fijaos que Sherlock usa la expresión *altamente improbable*. Esto indica que no está usando la deducción. Y no es inducción porque no está yendo de lo específico a lo general, en cierto modo está yendo de un conjunto de específicos (la observación que hace del apartamento de la víctima) a otro específico (declararlo asesinato en vez de suicidio). Abducción, querido Watson.

Argumentos

Cuando se ofrecen pruebas que respaldan un enunciado, estos enunciados combinados adquieren un estatus especial, lo que los lógicos llaman un argumento. Aquí *argumentar* no significa disputa o desacuerdo con alguien; significa un sistema lógico formal de enunciados. Los argumentos tienen dos partes: prueba y conclusión. La prueba puede ser uno o más enunciados o premisas. (Un enunciado sin prueba o sin conclusión no es un argumento en este sentido de la palabra).

Los argumentos establecen un sistema. Frecuentemente empezamos con la conclusión, sé que esto suena al contrario, pero es como hablamos típicamente; establecemos una conclusión y *después* revelamos la prueba.

Conclusión: Jacques hace trampa en el billar.

Prueba (o premisa): Cuando estabas de espaldas, lo vi mover la bola antes de darle.

El razonamiento deductivo sigue el proceso en la dirección opuesta.

Premisa: Cuando estabas de espaldas, lo vi mover la bola antes de darle.
Conclusión: Jacques hace trampa en el billar.

Esto está íntimamente relacionado con cómo hablan los científicos sobre los resultados de los experimentos, que son un tipo de argumento, de nuevo en dos partes.

Hipótesis = H
Implicación = I
H: No hay cisnes negros.
I: Si H es verdad, entonces ni yo ni nadie veremos jamás un cisne negro.
Pero *I* no es verdad. Mi tío Ernie vio un cisne negro, y después me llevó a verlo a mí también.
Por tanto, rechazo H.

Un argumento deductivo

La teoría microbiana de la enfermedad fue descubierta aplicando la deducción[180]. Ignaz Semmelweis era un médico húngaro que llevó a cabo un conjunto de experimentos (doce años antes de las investigaciones de Pasteur sobre gérmenes y bacterias) para determinar qué estaba causando la alta tasa de mortalidad en el ala de maternidad del Hospital General de Viena. El método científico no estaba bien establecido por aquel entonces, pero sus observaciones y manipulaciones sistemáticas ayudaron no solo a descubrir el culpable, sino también a avanzar en el conocimiento científico. Sus experimentos son un modelo de lógica deductiva y razonamiento científico.

En el contexto había una especie de condición de control: el Hospital General de Viena tenía dos alas de maternidad adyacentes, la primera (con tasa de mortalidad alta) y la segunda

(con tasa de mortalidad baja). Nadie podía descifrar por qué la tasa de muertes de bebés y de madres era mucho más alta en un ala que en la otra.

Una explicación ofrecida por una comisión de investigación fue que la configuración de la primera ala promovía angustia psicológica: cada vez que se llamaba a un sacerdote para darle la extremaunción a una mujer moribunda, tenía que pasar por las camas de maternidad de la primera ala para llegar a ella; era precedido por una enfermera tocando una campana. Se creía que esta combinación aterrorizaba a las mujeres que daban a luz y que, por tanto, las hacía más propensas a esta «fiebre puerperal». El sacerdote no tenía que pasar entre las madres parturientas de la segunda ala cuando daba la extremaunción porque tenía acceso directo a la habitación donde se encontrase la mujer moribunda.

Semmelweis propuso una hipótesis y una implicación que describieron un experimento:

H: La presencia de la campana que suena y del sacerdote incrementa la probabilidad de infección.

I: Si la campana y el sacerdote no están presentes, la infección no se incrementa.

Semmelweis convenció al sacerdote de que diera un rodeo para evitar pasar entre las madres parturientas de la primera ala y a la enfermera de que dejase de tocar la campana. La tasa de mortalidad no menguó.

I no es cierto.

Por tanto, H es falso.

Rechazamos la hipótesis después de una cauta experimentación.

Semmelweis contemplaba otras hipótesis. No era el hacinamiento, porque, de hecho, la segunda ala era la más abarrotada. No era la temperatura o la humedad, porque eran las mismas en ambas alas. Como frecuentemente ocurre en los descubrimientos científicos, un suceso de azar, puramente fortuito, llevó a una intuición. Un buen amigo de Semmelweis se

cortó accidentalmente con el bisturí de un estudiante que acababa de realizar una autopsia. El amigo enfermó gravemente y la posterior autopsia reveló algunos de los mismos síntomas de infección que se manifestaban en las mujeres que morían durante el parto. Semmelweis se preguntó si habría una conexión entre las partículas o sustancias químicas presentes en los cadáveres y la propagación de la enfermedad. Otra diferencia entre las dos alas que había pasado desapercibida hasta ahora cobró relevancia: el servicio de la primera eran estudiantes de medicina que frecuentemente estaban realizando autopsias o disecciones de cadáveres cuando se les reclamaba para atender un parto; el servicio de la segunda eran comadronas que no tenían otras tareas. Lavarse las manos no era una práctica común entre los doctores, por lo que Semmelweis propuso lo siguiente:

H: La presencia de contaminantes cadavéricos en las manos de los doctores aumenta la probabilidad de infección.

I: Si se neutralizan los contaminantes, la infección no aumentará.

Por supuesto, una *I* alternativa también era posible: si los trabajadores en ambas alas fueran intercambiados (si las comadronas atendieran en la primera ala, y los estudiantes de medicina en la segunda) la infección disminuiría. Esta era una implicación válida también, pero había dos razones por las que intercambiar a los trabajadores no era tan buena idea como hacer que los doctores se lavaran las manos. Primero, si la hipótesis era realmente cierta, la tasa de muertes en el hospital seguiría siendo la misma, todo lo que Semmelweis habría hecho sería cambiar las muertes de un ala a otra. Segundo, cuando no estuvieran atendiendo partos, los doctores aún tenían que trabajar en sus laboratorios de la primera ala y entonces habría un incremento en el retraso de ambos grupos de trabajadores para llegar hasta las madres parturientas, lo que podría contribuir a producir más muertes. Hacer que los doctores se lavaran las manos tenía la ventaja de que, si funcionaba, disminuiría la tasa de muertes de todo el hospital.

Semmelweis llevó a cabo el experimento pidiendo a los doctores que se desinfectaran las manos con una solución que contenía cloro. La tasa de mortalidad en la primera ala se redujo de un 18% a menos de un 2%.

---

177 S. Reardon (1 de julio de 2015), «US vaccine researcher sentenced to prison for fraud», *Nature,* 523, p. 138.

178 A. J. Wakefield *et al.* (28 de febrero de 1998), «retracted: Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children», *Lancet,* 351(9103), 637-641. http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(97)11096-0/abstract.

J. F. Burns (25 de mayo de 2010), «British medical council bars doctor who linked vaccine with autism», *New York Times,* p. A4. http://www.nytimes.com/2010/05/25/health/policy/25autism.html.

Associated Press (6 de enero de 2011), «Study linking vaccine to autism is called fraud», *New York Times,* http://query.nytimes.com/gst/fullpage.html?res=9C02E7DC1E3BF935A35752C0A9679D8B63.

T. S. Rao y C. Andrade (2011), «The MMR vaccine and autism: sensation, refutation, retraction, and fraud», *Indian Journal of Psychiatry,* 53(2), 95-96.

179 Tomado de S. Thompson (2010), «The blind banker», *Sherlock* (serie de televisión, emitida por primera vez el 31 de octubre de 2010).

180 La primera vez que encontré esta historia fue en C. Hempel (1966), *Philosophy of Natural Science,* Englewood Cliffs, NJ, Prentice-Hall.

FALACIAS LÓGICAS

Correlaciones ilusorias

El cerebro es un gigantesco detector de patrones y procura extraer orden y estructura de lo que a menudo parecen ser configuraciones aleatorias. Vemos Orión en el cielo nocturno no porque las estrellas estén organizadas así, sino porque nuestros cerebros son capaces de proyectar patrones sobre lo aleatorio.

Cuando un amigo nos llama justo cuando estamos pensando en él, ese tipo de coincidencia es tan sorprendente que nuestro cerebro lo registra. Lo que no registra tan bien son todas aquellas veces en las que *no estábamos* pensando en nadie y nos llamaron. Podemos visualizarlo como una de esas tablas de doble entrada de la primera parte. Suponed que es una semana particularmente increíble, llena de coincidencias (se os cruza un gato negro mientras camináis por un vertedero lleno de espejos rotos, subís hasta la decimotercera planta de un edificio y os encontráis con que están poniendo *Viernes trece* en una televisión). Digamos que recibís veinte llamadas esa semana y dos de ellas eran de amigos perdidos hace mucho sobre los que no habíais pensado en un tiempo, pero llamaron diez minutos después de pensar en ellos. Esa es la fila superior de la tabla: veinte llamadas, de las cuales dos fueron evocadas usando una señalización extrasensorial y dieciocho que no. ¡Un momento! Tenemos que rellenar la fila inferior de la tabla[181]: ¿cuántas veces estuvisteis pensando en personas y *no* llamaron, y —esta es mi favorita— cuántas veces *no* estabais pensando en alguien y os llamaron?

		¿Estaba pensando en ellos justo antes?		
		SÍ	NO	
Alguien llamó	SÍ	2	18	20
	NO	50	930	980
		52	948	1.000

Para rellenar el resto de la tabla, digamos que durante la semana 52 veces estabais pensando en alguien, y 930 veces durante la semana no estabais pensando en alguien. (Esto último es solo una suposición loca, pero si dividimos las 168 horas de la semana en fragmentos de diez minutos, eso son unos 980 pensamientos totales, y ya sabemos que 50 de esos fueron sobre personas que no os llamaron, dejamos 930 pensamientos sobre asuntos que no son personas; esto probablemente es una infraestimación, pero la cuestión queda clara con cualquier cifra razonable que queráis poner ahí, intentadlo vosotros mismos.)

El cerebro sólo detecta el recuadro superior izquierdo e ignora los otros tres, muy en detrimento del pensamiento lógico (y en favor del pensamiento mágico). Antes de que reservéis un viaje a Las Vegas para jugar a la ruleta, vamos a hacer las cuentas. ¿Cuál es la probabilidad de que alguien llame *dado* que acabáis de pensar en esa persona? Solo dos de cincuenta y dos o el 4%. Eso es, un 4% de las veces cuando estamos pensando en alguien nos llama. No parece tan impactante.

¿Qué podría explicar ese 4% de las veces en las que sucede esta coincidencia? Un físico podría simplemente apelar a los 1.000 eventos de la tabla de doble entrada y señalar que sólo dos de ellos (dos décimos del 1%) parecen ser «raros» y que deberíamos esperar eso por azar. Un psicólogo social podría preguntarse si hubo algún evento externo que causase que tanto nuestro amigo como nosotros pensásemos en el otro y que por consiguiente nos impulsara a llamar. Leísteis acerca de los ataques terroristas en París el 13 de noviembre de 2015. En

algún lugar en un rincón de nuestra mente, recordamos que nosotros y una amiga de la universidad siempre hablábamos de ir a París. Ella nos llama y nos sorprendemos tanto al escucharla que olvidamos la conexión con París, pero ella está reaccionando al mismo suceso y por eso cogió el teléfono.

Si esto os recuerda la historia previa de los gemelos criados por separado, es porque debería hacerlo. La correlación ilusoria es la explicación estándar ofrecida por los genetistas comportamentales para la confluencia extraña de comportamientos, como que ambos gemelos se rasquen la cabeza con el dedo corazón o como que ambos enrollen los bolígrafos con cinta adhesiva para mejorar el agarre. Estamos fascinados por los contenidos de la casilla superior izquierda de la tabla de doble entrada, obsesionados con todas las cosas que ambos gemelos hacen. Tendemos a ignorar todas las cosas que un gemelo hace y el otro no.

El contexto de las probabilidades

Después de la llamada de teléfono de nuestra antigua amiga de la universidad, decidimos ir a París de vacaciones una semana el próximo verano. Mientras estamos frente a la *Mona Lisa,* oímos una voz familiar y al volver la cabeza encontramos a nuestro antiguo compañero de habitación de la universidad, Justin, al cual no vemos desde hace años. «¡No me lo puedo creer!», dice Justin. «¡Lo sé!», dices tú. «¡Cuál es la probabilidad de cruzarme contigo aquí en París, parado ante la *Mona Lisa*! ¡Debe ser de una entre millones!».

Sí, la probabilidad de cruzarnos con Justin delante de la *Mona Lisa* seguramente es de una entre millones (sería difícil de calcular precisamente, aunque cualquier cálculo que hagamos dejaría claro que es muy pequeña). Pero dar ese contexto a la probabilidad es erróneo. Vamos a dar un paso atrás. ¿Qué pasaría si no nos hubiéramos cruzado a Justin justo mientras estábamos delante de la *Mona Lisa,* sino delante de la *Venus de Milo* o en *les toilettes,* o incluso mientras estábamos entrando?

¿Y si nos hubiéramos cruzado con Justin en el hotel, en una cafetería o en la Torre Eiffel? Nos hubiéramos quedado igual de sorprendidos. Para el caso, olvidemos a Justin, si nos hubiéramos encontrado a *cualquier conocido* durante esas vacaciones, en *cualquier lugar de París,* estaríamos igual de sorprendidos. ¿Por qué limitarlo a las vacaciones en París? Podría haber sido en un viaje de negocios a Madrid, mientras hacíamos escala en Cleveland, o en un balneario en Tucson. Vamos a contextualizar la probabilidad de la siguiente manera: en algún momento de nuestra vida adulta, nos cruzaremos con alguien conocido en algún lugar donde no esperaríamos encontrarlo. Está claro que la probabilidad de que eso ocurra es bastante alta. Pero el cerebro no piensa automáticamente así, la ciencia cognitiva nos ha enseñado cuán necesario es que nos entrenemos para evitar el pensamiento defectuoso.

El contexto del riesgo

Un problema relacionado con el contexto de las probabilidades es no conseguir dotar de un contexto lógico a los riesgos. Incluso teniendo en cuenta la catástrofe de los aviones de los ataques del 11 de septiembre en Estados Unidos, viajar en avión continuó (y continúa siendo) el método más seguro de transporte, seguido de cerca por el transporte ferroviario.

Las probabilidades de morir en un vuelo comercial o en un viaje en tren son próximas a cero. Aun así, justo después del 11 de septiembre, muchos viajeros estadounidenses evitaron los aviones y en su lugar utilizaron las carreteras. Las muertes por accidentes de automóvil sufrieron un incremento drástico. La gente siguió su intuición emocional en lugar de una respuesta lógica, ajenos al aumento del riesgo. La *tasa* de accidentes de vehículos no subió por encima de la línea de base, pero la suma de personas que murieron en todos los accidentes relacionados con el transporte se incrementó a medida que más personas escogían una manera menos segura de viajar[182].

Podríamos sacar una estadística como esta:

Murió más gente en accidentes de avión en 2014 que en 1960.

De esto, podríamos concluir que viajar en avión se ha vuelto mucho menos seguro[183]. La estadística es correcta, pero no es la estadística relevante. Si pretendemos dilucidar cuán seguro es viajar en avión, considerar la cantidad total de muertes no nos lo va a aclarar. Es necesario considerar la *tasa* de muertes, las muertes por kilómetros volados, o las muertes por vuelo, o algo que iguale el punto de referencia. En 1960 no había la misma cantidad de vuelos ni por asomo, pero eran más peligrosos[184].

Con una lógica similar se puede decir que mueren más personas en las carreteras entre las cinco y las siete de la tarde que entre las dos y las cuatro de la mañana, por lo que habría que evitar conducir entre las cinco y las siete. Pero el hecho claro es que hay mucha más gente conduciendo entre las cinco y las siete, hay que considerar la *tasa* de muertes (por kilómetros o por viaje o por coche), no la cifra bruta. Al hacerlo nos daremos cuenta de que conducir por la tarde es más seguro (en parte porque es más probable que la gente que hay en la carretera entre las dos y las cuatro de la mañana esté ebria o falta de sueño).

Después de los ataques de París del 13 de noviembre del 2015, la CNN informó que al menos uno de los atacantes había entrado en la Unión Europea como refugiado, con un telón de fondo de un clima creciente de sentimientos antirrefugiados en Europa. Los activistas antirrefugiados habían estado pidiendo un control más estricto de las fronteras. Se trata de un conflicto social y político y no es mi intención tomar partido, pero los datos pueden informar la toma de decisiones. Cerrar las fronteras por completo a los emigrantes y los refugiados podría haber frustrado los ataques que segaron unas 130 vidas. Denegar la entrada a un millón de emigrantes provenientes de regiones devastadas por la guerra como Siria y Afganistán podría, con seguridad, haber costado miles de vidas, muchas más que los 130 que murieron en los ataques. Hay otros riesgos en ambos cursos de acción y otras consideraciones. Pero para alguien que no está pensando mediante la lógica de los números, un titular como «Uno de los

atacantes era un refugiado» exacerba las emociones sobre el sentimiento antiinmigrantes, sin tener en cuenta la cantidad de vidas que salvan las políticas de inmigración. La mentira que los terroristas quieren que creamos es que estamos en un peligro enorme e inminente.

Sacar de contexto es algo frecuentemente utilizado por los comerciales para persuadirnos a comprar sus productos. Supongamos que recibimos un correo electrónico de una empresa de seguridad para el hogar con este tono: «El 90% de los robos en casas se resuelve gracias a grabaciones proporcionadas por el dueño de la casa». Suena tan empírico. Tan científico.

Empecemos revisando la verosimilitud. Olvidemos la segunda parte de la frase, sobre el vídeo, y limitémonos a la primera parte: «El 90% de los robos en casas se resuelve...» ¿Suena razonable? Sin buscar las estadísticas reales, simplemente usando nuestro conocimiento del mundo real, suena dudoso que el 90% de los robos en casas se resuelvan. Eso sería una tasa de éxito altísima para el cuerpo de policía. Vayamos a Internet. Una página del FBI informa de que alrededor del 30% de los casos de robo se «aclaran», en el sentido de que se resuelven[185].

Así que podemos rechazar la frase inicial como altamente improbable. Afrmaba que el 90% de los robos en casas se resuelven con grabaciones proporcionadas por el dueño de la casa. Pero eso no puede ser verdad, implicaría que más del 90% de los robos en casas se resuelven, porque algunos seguro que se resuelven sin grabaciones. Con mayor probabilidad, a lo que se refiere la empresa es que el 90% de los robos resueltos lo son por grabaciones proporcionadas por el dueño de la casa.

¿No es lo mismo?

No, porque la muestra es diferente. En el primer caso, nos fijamos en todos los robos cometidos. En el segundo caso, únicamente en los que se resuelven, un número mucho menor.

Aquí está visualmente:

Todos los robos de un barrio[186]



Robos resueltos de un barrio (utilizando por la cifra del 30% obtenida anteriormente).



¿Significa esto que si tengo una cámara de vídeo hay un 90% de probabilidades de que la policía fuese capaz de resolver un robo en mi casa?

¡No!

Lo único que sabemos es que *si* un robo se resuelve, hay un 90% de probabilidades de que la policía se ayudara de una grabación. Si estamos pensando que tenemos información suficiente para responder a la pregunta que verdaderamente nos interesa (cuál es la probabilidad de que la policía sea capaz de resolver un robo en nuestra casa si compramos un sistema de seguridad frente a si no lo compramos), nos equivocamos, necesitamos hacer una tabla de doble entrada como las de la primera parte, pero al empezar nos damos cuenta de que solo tenemos información de una de las filas. Sabemos qué porcentaje de crímenes resueltos tenían grabaciones, pero para completar la tabla de doble entrada, también necesitamos saber qué proporción de los crímenes sin resolver tenían grabaciones de la casa (o, dicho de otra manera, qué proporción de las grabaciones tomadas resultaron en crímenes sin resolver).

Recordad, P(robos resueltos | grabaciones de la casa) ≠ P(grabaciones de la casa | robos resueltos).

Sacar de contexto estos datos pretende estimular nuestras emociones y hacer que compremos un producto que podría no tener en absoluto el resultado que se busca.

Persistencia de las creencias

Una característica peculiar de la cognición humana es que una vez que formamos una creencia o aceptamos una afirmación, es muy difícil que nos deshagamos de ella, incluso frente a una evidencia abrumadora o a pruebas científicas que demuestren lo contrario. La investigación nos dice que deberíamos comer una dieta baja en grasa y alta en carbohidratos y lo hacemos.

Nuevas investigaciones restan apoyo al descubrimiento anterior de manera bastante convincente, pero aun así somos reacios a cambiar nuestros hábitos alimenticios. ¿Por qué? Porque al adquirir la información nueva, tendemos a construir historias internas que nos ayudan a asimilar el conocimiento. Nos decimos «comer grasas me pondrá gordo», «así que la dieta baja en grasa tiene mucho sentido». Leemos sobre un hombre condenado por un homicidio espeluznante. Vemos su foto en un periódico y pensamos que podemos descifrar los ojos pequeños y brillantes y la implacable mandíbula de un asesino a sangre fría. Nos convencemos de que «parece un asesino». Con las cejas arqueadas, la boca relajada, parece carecer de remordimientos. Luego, cuando se le absuelve sobre la base de pruebas exculpatorias, no podemos desprendernos de la sensación de que aunque no haya cometido *este* asesinato, tiene que haber cometido otro. Si no, no parecería tan culpable.

En un famoso experimento psicológico[187] a los participantes se les mostraron fotos de miembros del sexo opuesto, mientras se suponía que estaban conectados a un equipo de monitorización fisiológica que indicaba sus niveles de excitación. En realidad, no estaban conectados al equipo en absoluto, sino bajo el control del experimentador. El experimentador hizo

comentarios falsos a los participantes para hacerles creer que habían mostrado especial atracción por una persona de las fotos en concreto, más que hacia las demás. Cuando el experimento acabó, se mostró a los participantes que las «reacciones» de su propio cuerpo eran en realidad grabaciones prefabricadas. La sorpresa es que el experimentador permitió que se llevaran una foto a casa. Lógicamente hubieran debido elegir la foto que les pareciese más atractiva en ese momento, pues la prueba de que les gustara una foto en particular había sido completamente desacreditada. Pero los participantes tendían a escoger la foto que coincidía con su creencia inicial. Los experimentadores mostraron que el efecto era causado por el tipo de autopersuasión descrito previamente.

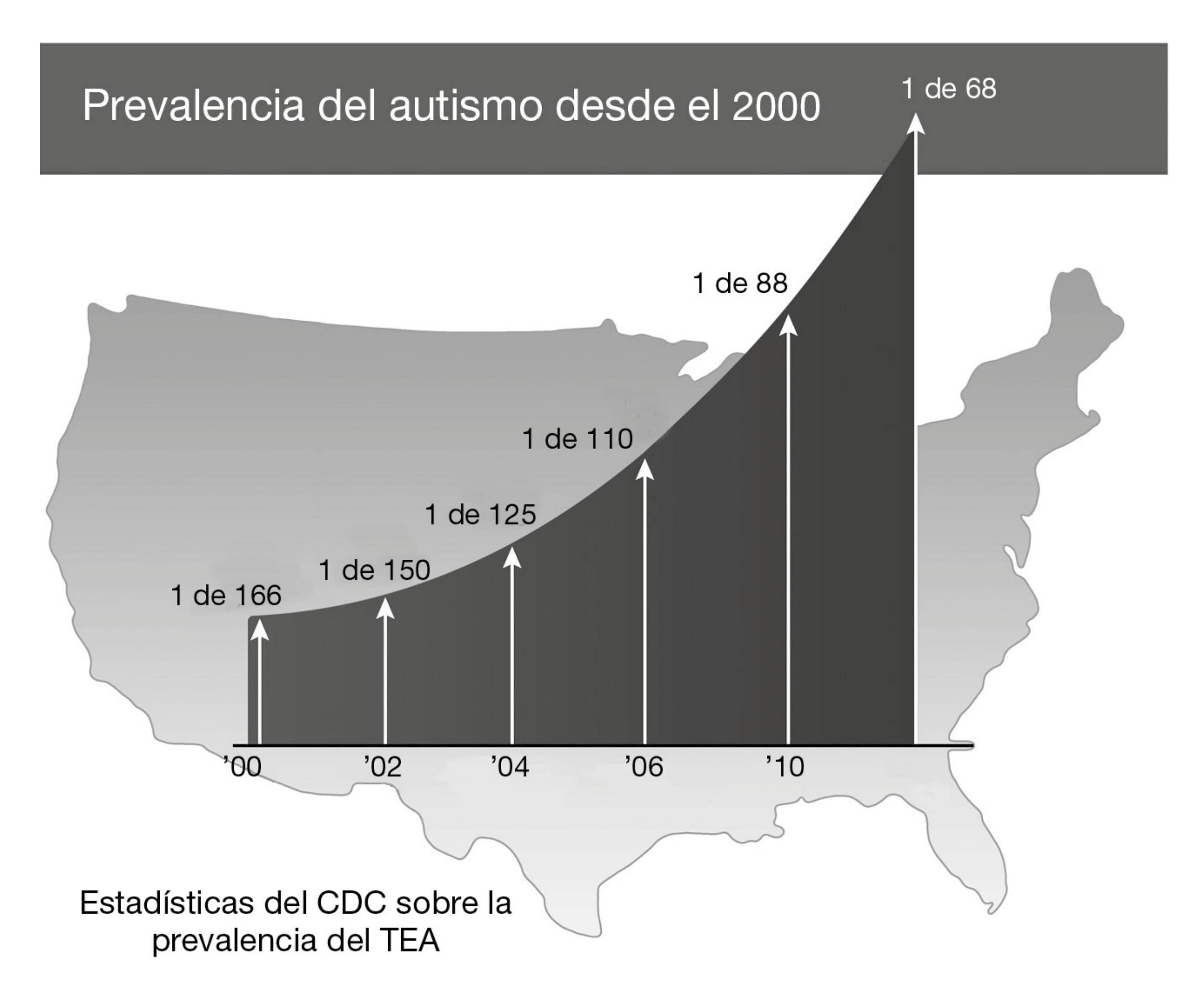
Autismo y vacunas: cuatro dificultades del razonamiento

La historia del autismo y las vacunas atañe a cuatro dificultades distintas del pensamiento crítico: la correlación ilusoria, la persistencia de la creencia, la persuasión por asociación y la falacia lógica que vimos antes, *post hoc ergo propter hoc* (vagamente traducido, significa «como esto pasó después de eso, eso debe haber causado esto»).

Entre 1990 y 2010, el número de niños diagnosticados con trastornos del espectro autista (TEA) se sextuplicó[188], y aumentó más del doble en los últimos diez años (véase el gráfico de la página siguiente). La prevalencia del autismo se ha incrementado exponencialmente desde 1970 hasta ahora.

Casi todo el aumento[189] ha sido justificado por tres factores: el aumento de la concienciación sobre el autismo (más padres están alerta y llevan a evaluar a sus hijos, los profesionales están más dispuestos a diagnosticarlo); la ampliación de las definiciones, lo que incluye más casos; y el hecho de que las personas tienen hijos más tarde (la edad parental avanzada se correlaciona con la probabilidad de tener hijos con autismo y muchos otros trastornos).

Si permitimos que Internet guíe nuestras opiniones sobre por qué ha aumentado el autismo[190], se nos presentará una panoplia de causas malignas: los OGM (organismos genéticamente modificados), el azúcar refinado, las vacunas infantiles, los glifosatos, las wifi y la proximidad a las autopistas. ¿Qué tiene que hacer un ciudadano preocupado? Estaría muy bien si un experto interviniera. *Voilà,* ¡un científico del MIT viene al rescate! La Dra. Stephanie Seneff llegó a los titulares en 2015 cuando informó de una conexión entre el aumento del uso de glifosato, un componente activo en el herbicida Roundup, y el aumento del autismo. Correcto, dos cosas aumentan, como los piratas y el calentamiento global, así que debe haber una conexión causal, ¿no?



¿*Post hoc, ergo propter hoc,* hay alguien ahí?

La Dra. Seneff es una científica computacional sin formación en agricultura, genética o epidemiología. Pero es una *científica* en

el venerable MIT, así que muchas personas asumen erróneamente que su pericia va más allá de su formación. Además, formula su argumento con lenguaje científico[191], dándole un verdadero lustre pseudocientífico de contraconocimiento:

1. El glifosato interrumpe la ruta del ácido shikímico en las plantas.
2. La ruta del ácido shikímico permite crear aminoácidos a las plantas.
3. Cuando se interrumpe la ruta, la planta muere.

Seneff concede que las células humanas no tienen una ruta del ácido shikímico, pero continúa:

4. Tenemos millones de bacterias en nuestro intestino (flora intestinal).
5. Esas bacterias sí que tienen ruta del ácido shikímico.
6. Cuando el glifosato entra en nuestro sistema, perturba la digestión y nuestro sistema inmune.
7. El glifosato también puede inhibir la función del hígado en humanos.

Seguramente os estáis preguntando qué tiene todo esto que ver con el TEA, y con razón. Seneff plantea la hipótesis (sin citar ninguna prueba) de la relación entre la prevalencia de los problemas digestivos y la disfunción del sistema inmune, pero eso no tiene nada que ver con el TEA.

Otras búsquedas para encontrar una explicación al aumento de las tasas de autismo han señalado a la vacuna de la SPR (sarampión-paperas-rubeola) y el compuesto antiséptico, antifúngico timerosal (tiomersal) que contiene. El timerosal es un derivado del mercurio, y la cantidad que contiene la vacuna es típicamente un cuarentavo de lo que la Organización Mundial de la Salud (OMS) considera tolerable por día. Obsérvese que las directivas de la OMS se expresan en cantidades por día y que la vacuna solo se toma una vez.

Aunque no hay pruebas de que el timerosal esté relacionado con el autismo[192], en 1992 se eliminó de las vacunas en Dinamarca y Suecia y 1999 en Estados Unidos, como «medida de precaución». La tasa de autismo ha seguido incrementándose rápidamente, incluso con la retirada del agente. La correlación

ilusoria (como en los piratas y el calentamiento global) es que la vacuna del SPR se suele aplicar entre los doce y los quince meses de edad, y si un niño tiene autismo, lo más pronto que normalmente se diagnostica es entre los 18 y los 24 meses de edad. Los padres tendían a fijarse en la casilla superior izquierda de la tabla de doble entrada, la cantidad de veces que un niño ha recibido la vacuna y posteriormente ha sido diagnosticado con autismo, sin considerar cuántos niños que no fueron vacunados aun así fueron diagnosticados de autismo, o cuántos millones de niños fueron vacunados y no desarrollaron autismo.

Para empeorar las cosas, un médico ahora desacreditado, Andrew Wakefield, publicó un artículo científico en 1998 donde afirmaba la existencia de una conexión. El *British Medical Journal* declaró que su trabajo era fraudulento y seis años después la revista que originalmente lo publicó, el *Lancet,* lo retiró. Su licencia médica fue revocada. Wakefield era cirujano, no un experto en epidemiología, toxicología, genética, neurología o ninguna otra especialidad que pudiera haberlo cualificado como experto en autismo.

Post hoc, ergo propter hoc provocó que la gente creyera que la correlación implicaba causación. La correlación ilusoria hizo que se centraran solo en la coincidencia de que algunas personas que desarrollaron autismo también se habían vacunado. El testimonio de una científica computacional y un médico hicieron que la gente se convenciese por asociación. La persistencia de la creencia provocó que la gente que inicialmente había creído en la conexión se aferrara a sus creencias incluso después de que las pruebas hubieran sido eliminadas.

Los padres continúan culpando a la vacuna por el autismo y muchos han dejado de vacunar a sus hijos. Esto llevó a varios brotes de sarampión en distintos lugares del mundo. Todo por una relación espuria y la incapacidad de muchas personas para distinguir entre correlación y causación, y la incapacidad de formar creencias basadas en lo que hoy día son pruebas científicas inapelables.

---

181 Utilizo la semana de 168 horas (7 días × 24 horas por día) para dar cuenta de los pensamientos que podríamos tener mientras dormimos y de las personas que podrían llamarnos y despertarnos mientras dormimos. Es cierto que podríamos restar ocho horas de sueño por noche (o las que sean) y utilizar únicamente las 112 horas de vigilia restantes para obtener una probabilidad diferente, pero eso no altera la conclusión.

182 Retrospectivamente ese cambio de opción fue estúpido, pero en su momento puede haber sido una elección racional. Cuatro aviones secuestrados por suicidas constituyen algo sin precedentes en la historia de la aviación. Cuando nos enfrentamos a un cambio sustancial en el mundo, a menudo la mejor opción es pensar de modo bayesiano: actualizar nuestro modo de entender el mundo, dejar de confiar en las estadísticas antiguas y buscar alternativas.

183 Basado en Huff, *op. cit.,* p. 79.

184 Véanse, por ejemplo, C. Iolan, T. Patterson y A. Johnson (28 de julio de 2014), «Is 2014 the deadliest year for fights? Not even close», CNN, http://www.cnn.com/interactive/2014/07/travel/aviation-data/; y N. Evershed (28 de marzo de 2015), «Aircraft accident rates at historic low despite high-profile plane crashes», *Guardian,* http://www.theguardian.com/world/datablog/2014/dec/29/aircraft-accident-rates-at-historic-low-despite-high-profile-plane-crashes.

185 http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2011/crime-in-the-u.s./2011/clearances.

186 Imagen tomada de http://www.clipartof.com/portfolio/bestvector/illustration/retro-black-and-white-criminal-carrying-a-flashlight-and-box-210461.html.

187 R. E. Nisbett y S. Valins (1972), «Perceiving the causes of one's own behavior», en D. E. Kanouse *et al.* (eds.), *Attribution: perceiving the causes of behavior,* Morristown, NJ, General Learning Press, pp. 63-78.

Y S. Valins (2007), «Persistent effects of information about internal reactions: ineffectiveness of debriefing», en H. London y R. E. Nisbett (eds.), *Thought and Feeling: the cognitive alteration of feeling states,* Chicago, IL, Aldine Transaction.

188 «What is causing the increase in autism prevalence», *Autism Speaks Offcial Blog,* 22 de octubre de 2010, http://blog.autismspeaks.org/2010/10/22/got-questions-answers-to-your-questions-from-the-autism-speaks%E2%80%99-science-staff-2/.

189 *Ibid*.

190 A. Suresh (13 de octubre de 2015), «Autism increase mystery solved: no, it's not vaccines, GMOs, glyphosate —or organic foods. Genetic Literacy

Project», http://www.geneticliteracyproject.org/2015/10/13/autism-increase-mystery-solved-no-its-not-vaccines-gmos-glyphosate-or-organic-foods/.

191 A. Kase (11 de mayo de 2015), «MIT scientist uncovers link between glyphosate, GMOs and the autism epidemic», *Reset.me,* http://reset.me/story/mit-scientist-uncovers-link-between-glyphosate-gmos-and-the-autism-epidemic/.

192 H. Honda, Y. Shimizu y M. Rutter (2005), «No effect of MMR withdrawal on the incidence of autism: a total population study», *Journal of Child Psychology and Psychiatry,* 46(6), 572-579. http://1796kotok.com/pdfs/MMR_withdrawal.pdf, y muchas otras fuentes.

S. Reardon (2015), «US vaccine researcher sentenced to prison for fraud», *Nature,* 523(7559), p. 138.

SABER LO QUE NO SE SABE

... como sabemos, hay hechos conocidos que conocemos. Hay cosas que sabemos que sabemos. También sabemos que hay hechos conocidos que desconocemos. Es decir, sabemos que hay ciertas cosas que no sabemos. Pero también hay hechos desconocidos que desconocemos, los que no sabemos que no sabemos[193].

Donald Rumsfeld,
secretario de Defensa de los Estados Unidos.

Se trata muy claramente de un lenguaje tortuoso y el significado de la oración queda oscurecido por esa razón. No hay motivo para el empleo repetitivo de la misma palabra y el secretario hubiera sido más claro si hubiera dicho: «Hay cosas que sabemos, cosas que somos conscientes de que no sabemos y algunas cosas de las que ni siquiera tenemos conciencia de que no las sabemos». Sin duda, hay una cuarta posibilidad, cosas que sabemos y que no somos conscientes de que las sabemos. Probablemente lo habréis experimentado alguna vez, alguien os hace una pregunta y la respondéis, aunque os digáis a vosotros mismos «ni siquiera sé cómo sabía eso».

De cualquier modo, la cuestión fundamental es sólida: lo que realmente puede dañarnos y provocarnos serios malestares e inconvenientes son aquellas cosas que creemos saber pero no sabemos (según el epigrama de Mark Twain/Josh Billings del comienzo del libro) y aquellas cosas de las que ni siquiera tenemos conciencia de que son extremadamente pertinentes para la decisión que hemos de tomar (la incógnita de lo desconocido). Plantear correctamente una pregunta científica supone tener en cuenta lo que sabemos y lo que no sabemos. Una hipótesis científica correctamente formulada es *falsable:* hay pasos que podemos dar, al menos en teoría, para poner a prueba el

verdadero estado del mundo, para determinar si la hipótesis es verdadera o no. En la práctica esto significa prever las explicaciones alternativas y diseñar un experimento de modo que puedan excluirse las alternativas.

Si ensayamos un fármaco nuevo en dos grupos de personas, las condiciones experimentales han de ser las mismas, a fin de poder concluir que el fármaco A es mejor que el fármaco B. Si todas las personas del grupo A toman su fármaco en una habitación con ventanas y buenas vistas y las personas del grupo B lo toman en un apestoso laboratorio del sótano, estará presente un factor de distorsión que no nos permitirá concluir que la diferencia, si es que encontramos alguna, se debe únicamente a la medicación. El problema del sótano apestoso es una incógnita que conocemos. Si la medicina A funciona mejor que la medicina B es una incógnita que conocemos (por eso estamos realizando el experimento). La incógnita que desconocemos en este caso es si existe otro factor de distorsión potencial. Cabe que las personas con presión sanguínea alta respondan mejor al fármaco A en todos los casos y que las personas con presión sanguínea baja respondan mejor al fármaco B. Quizá el historial familiar sea importante; puede que la hora del día en que se toma la medicación suponga una diferencia. Una vez que identificamos un factor de distorsión potencial se traslada limpiamente desde la categoría de las incógnitas desconocidas a la de las incógnitas conocidas. Seguidamente podremos modificar el experimento o realizar más investigaciones que nos ayuden a aclararlo.

El secreto para diseñar buenos experimentos, o para evaluar los que ya se han realizado, en última instancia consiste en ser capaz de generar explicaciones alternativas. Se puede afirmar que *el* principal cometido de los científicos es descubrir las incógnitas de lo desconocido. Cuando un experimento produce resultados sorprendentes lo celebramos, porque supone una oportunidad de aprender algo que no conocíamos. El papel del científico de las películas de serie B, aferrado a su teoría favorita hasta su último aliento, no se aplica a ningún científico que conozca; los científicos de verdad saben que únicamente aprenden algo cuando las cosas no salen como creían.

En pocas palabras:

1. Hay algunas cosas que sabemos, como la distancia entre la Tierra y el Sol. Quizá no seamos capaces de dar la respuesta correcta sin consultarla, pero somos conscientes de que la respuesta es conocida. Esto es lo que Rummy llama *lo que sabemos que sabemos*.
2. Hay algunas cosas que no sabemos; por ejemplo, cómo las descargas neuronales producen sentimientos de alegría. Somos conscientes de que desconocemos la respuesta a esta pregunta. Esto es lo que Rummy llama *lo que sabemos que no sabemos*.
3. Hay algunas cosas que sabemos, pero no somos conscientes de que las sabemos, o nos olvidamos de que las sabemos. ¿Cuál era el segundo apellido de nuestra abuela? ¿Quién se sentaba a nuestro lado en tercero de Primaria? Si un indicio correcto de recuerdo nos ayuda a recordar algo, nos damos cuenta de que lo sabíamos, aunque no éramos conscientes de antemano. Pese a que Rummy no lo menciona, esto es una *incógnita conocida,* lo que no sabemos que sabemos.
4. Hay algunas cosas que no sabemos, de las que ni siquiera somos conscientes que no las sabemos. Si compramos una casa, seguramente hemos pagado alguna inspección para asegurarnos del estado del tejado, los cimientos y si hay termitas o algún otro organismo que infecte la madera. Si nunca hemos oído hablar del radón y vuestro agente inmobiliario está más interesado por cerrar la venta que por la salud de vuestra familia, puede que no se os haya ocurrido indagarlo. Pero hay muchas casas que presentan niveles elevados de radón, un conocido carcinógeno. Esto sería una *incógnita desconocida,* lo que no sabemos que no sabemos (aunque al haber leído el párrafo anterior ya no lo es). Fijaos que ser consciente, o no, de una incógnita depende de vuestra pericia y vuestra experiencia. Un inspector de control de plagas os dejará claro que únicamente informa sobre lo que resulta visible, sabe que puede haber vicios ocultos en la casa, en zonas a las que no puede acceder. La naturaleza y amplitud de esos daños, si es que los hay, son desconocidas para él, pero es consciente de que podrían existir (una *incógnita conocida*). Si aceptáramos a ciegas su informe y supusiéramos que es completo, no seríamos conscientes de que podría haber otros daños (una *incógnita desconocida*).

Podemos aclarar las cuatro posibilidades del secretario Rumsfeld con una tabla de doble entrada[194]:

Lo que sabemos que sabemos: ESTUPENDO, GUÁRDALO EN EL BANCO	Lo que sabemos que no sabemos: NO ESTÁ MAL, PODRÍAMOS APRENDERLO
Lo que no sabemos que sabemos: UN REGALO	Lo que no sabemos que no sabemos: PELIGRO – ARRECIFES OCULTOS

Las incógnitas desconocidas son las más peligrosas. Algunos de los mayores desastres de la humanidad pueden remontarse hasta ellas. Cuando un puente se desploma, los países pierden guerras o al comprador de una casa le ejecutan la hipoteca, a menudo se debe a no haber considerado la posibilidad de que no lo sabían todo y han seguido adelante, creyendo que todas las posibilidades habían sido tenidas en cuenta. Uno de los principales objetivos de formar a una persona en un doctorado en Ciencias, en una licenciatura en Medicina o en Derecho, o en un MBA o en el mando de tropas es enseñarles a identificar y a pensar sistemáticamente sobre lo que no saben, a fin de convertir las incógnitas desconocidas en incógnitas conocidas.

Un último tipo al que no se refirió el secretario Rumsfeld son los conocimientos incorrectos, aquello que damos por sentado, pero no lo está. Creer enunciados falsos cae en esta categoría. Una de las principales causas de los resultados malos o incluso fatales es creer cosas que son mentira.

---

193 Defense.gov Transcripción de Noticias: DoDNews Briefing-Secretary Rumsfeld and Gen. Myers, United States Department of Defense (defense.gov).

194 Agradezco esta idea a Morris Olitsky.

PENSAMIENTO BAYESIANO EN LA CIENCIA Y EN LOS TRIBUNALES

Recordemos la noción de probabilidad bayesiana de la primera parte, mediante la que podemos modificar o actualizar nuestras creencias basándonos en datos nuevos según vamos obteniéndolos, o merced a la probabilidad previa de que algo sea verdad: la probabilidad de tener neumonía *dado* que se muestran ciertos síntomas, o la probabilidad de que una persona vote por un partido *dado* donde vive.

Según el enfoque bayesiano, asignamos una probabilidad subjetiva a la hipótesis (probabilidad *previa*) y después modificamos esa probabilidad a la luz de los datos recopilados (la probabilidad *posterior,* ya que es la que alcanzamos una vez realizado el experimento). Si teníamos razones para creer que la hipótesis era válida antes de ponerla a prueba, no necesitamos muchas pruebas para confirmarla. Si teníamos razones para creer que era inverosímil antes de ponerla a prueba, necesitaremos más pruebas.

Las afirmaciones inverosímiles según la perspectiva bayesiana requieren de pruebas más sólidas que las probables. Supongamos que un amigo nos dice que ha visto por la ventana algo que va volando. Podríamos albergar tres hipótesis: se trata de un petirrojo, se trata de un gorrión, se trata de un cerdo. Podemos asignar probabilidades a esas tres hipótesis. Ahora nuestro amigo nos muestra una foto de un cerdo volando fuera de la ventana. Nuestra creencia previa sobre que los cerdos vuelen es tan baja que la probabilidad posterior es todavía muy pequeña, pese a esa prueba. Probablemente lo que nos sucede ahora es que consideramos la hipótesis de que la foto está trucada o de que hay algún otro tipo de engaño. Si esto os trae a la cabeza las tablas de doble entrada y la probabilidad de que alguien padezca cáncer de mama cuando obtiene un resultado positivo en un test, estáis en la senda correcta: las tablas de doble entrada son simplemente un método para realizar cálculos bayesianos.

Los científicos deben establecer un umbral de aceptación más alto para las pruebas contrarias a las teorías estándar o establecidas que para las pruebas que son congruentes con lo que conocemos. Tras realizar miles de ensayos con éxito de un nuevo fármaco retroviral con ratones y monos, no nos sorprende descubrir que también funciona en humanos, estamos dispuestos a aceptar la evidencia cuando cumple los estándares habituales de validez. Podríamos convencernos con un único estudio que utilizara unos pocos cientos de participantes. Pero si alguien nos dice que el sida se puede curar sentándose durante tres días al pie de una pirámide, ya que esta hace de canal para el qi en nuestros chakras, vamos a necesitar pruebas más sólidas que un único experimento, ya que resulta inverosímil y jamás se ha demostrado algo así antes. Necesitaríamos que el resultado se reprodujera muchas veces y bajo muchas condiciones diferentes y, en última instancia, un metaanálisis.

El enfoque bayesiano no es el único procedimiento mediante el que los científicos abordan sucesos improbables. En su búsqueda del bosón de Higgs los científicos establecieron un umbral (mediante pruebas estadísticas convencionales, no bayesianas) 50.000 veces más restrictivo que el habitual, no porque el bosón de Higgs fuera improbable (se ha pronosticado su existencia durante décadas), sino porque el precio de equivocarse era muy alto (los experimentos a realizar son muy caros).

Quizá la mejor ilustración de la aplicación del teorema de Bayes sea un ejemplo de la ciencia forense. Uno de los pilares centrales de la ciencia forense[195] fue desarrollado por el médico y abogado francés Edmond Locard: todo contacto deja alguna traza. Locard afirmó que o bien el malhechor deja algún indicio en el escenario del crimen o se lo lleva consigo, sobre su persona o sus vestidos, indicios de dónde ha estado o de lo que ha hecho.

Supongamos que un criminal penetra en los establos[196] para drogar a un caballo la noche previa a una gran carrera. Dejará trazas de su presencia en el escenario del crimen, huellas de pasos, quizá un poco de piel, cabellos, fibras de tejido, etc. Las pruebas se han trasladado del criminal al escenario del crimen y, de modo semejante, el criminal recibe pruebas en el escenario del

crimen, suciedad, pelo de caballo, fibras de una manta, etc. que vienen del establo.

Ahora imaginemos que alguien es detenido al día siguiente. Se toman muestras de su ropa, sus manos, sus uñas e indicios de ese tipo, entre las pruebas que se recogen en el escenario del crimen. El fiscal del distrito quiere valorar la potencia de esas pruebas. Cabe que existan semejanzas porque el sospechoso es culpable. O tal vez sea inocente, pero estuvo en contacto con el culpable, también ese contacto dejaría alguna traza. O puede que el sospechoso estuviera inocentemente en otro establo, interactuara inocentemente con otro caballo y eso explicaría las semejanzas.

El empleo del teorema de Bayes nos permite combinar las probabilidades objetivas, como la probabilidad de que el ADN del sospechoso coincida con el ADN que se encontró en el escenario del crimen, con estimaciones personales, subjetivas, como la credibilidad de un testigo, la honestidad y el historial del agente de la oficina del forense que custodió las muestras de ADN. ¿Ha cometido el sospechoso algo parecido antes? o, si se trata de alguien que no sabe nada sobre carreras de caballos, ¿tiene alguna conexión con alguien relacionado con la carrera? ¿Tiene una coartada muy buena? Todos estos factores nos ayudan a determinar la probabilidad previa subjetiva de que el sospechoso sea culpable.

Si tomamos literalmente el presupuesto del sistema legal estadounidense[197] según el que alguien es inocente hasta que se demuestra que es culpable, entonces la probabilidad previa de que un sospechoso sea culpable es cero y cualquier prueba, por crítica que sea, no podrá elevar la probabilidad posterior por encima de cero, ya que siempre estaremos multiplicando por cero. Un modo más sensato de establecer la probabilidad previa de la inocencia de un sospechoso es considerar que toda la población presenta las mismas probabilidades. Por lo tanto, si el sospechoso fue detenido en una ciudad de 100.000 personas y los investigadores tienen motivos para creer que el criminal residía en esa ciudad, la probabilidad previa de que el sospechoso sea culpable es de 1 entre 100.000. Está claro que las pruebas pueden limitar el rango de la búsqueda entre la población, puede que se sepa, por ejemplo, que no había señales de que se entrara por la fuerza, de manera

que el sospechoso tiene que ser una de las cincuenta personas que tenían acceso a las instalaciones.

Nuestra hipótesis *a priori* es que el sospechoso es culpable con una probabilidad de 0,02 (una entre las cincuenta personas que tenían acceso). Ahora supongamos que el criminal y el caballo pelearon y se encontró sangre humana en el escenario del crimen. El equipo forense nos informa de que la probabilidad de que la sangre del sospechoso coincida con la que se encontró en el lugar del crimen es de 0,85. Elaboramos nuestra tabla de doble entrada como antes. Primero rellenamos la fila inferior de la tabla: el sospechoso tiene una probabilidad entre cincuenta de ser culpable (la columna *Culpable: Sí*) y una probabilidad de uno entre cuarenta y nueve de ser inocente. El laboratorio nos informa que hay una probabilidad de 0,85 de que coincida la sangre, así que introducimos los datos de la izquierda arriba: la probabilidad de que el sospechoso sea culpable *y* la sangre coincida. Eso significa que la celda inferior de la izquierda tiene que ser 0,15 (las probabilidades han de sumar uno). La coincidencia de sangre de 0,85 implica algo más: hay una probabilidad de 0,15 de que la sangre sea de otra persona distinta del sospechoso, lo que lo absolvería y significaría que no es culpable. Existe una probabilidad de 0,15 de que una de las personas de la columna de la derecha coincida, de modo que hemos de multiplicar 49 × 0,15 para obtener 7,35 en la celda superior de la derecha. Eso lo restamos de 49 a fin de obtener el valor de la celda inferior de la derecha.

		Sospechoso culpable		
		SÍ	NO	
Coincide la sangre	SÍ	0,85	7,35	8,2
	NO	0,15	41,65	41,8
		1	49	50

Ahora podemos calcular la información que queremos que valoren el juez y el jurado.

P(Culpable | Coincide) = 0,85/8,2 = 0,10
P(Inocente | Coincide) = 7,35/8,2 = 0,90

Dadas las pruebas, es aproximadamente nueve veces más probable que nuestro sospechoso sea inocente que sea culpable. Comenzamos con una probabilidad de 0,02 de que fuera culpable, en consecuencia la información nueva ha multiplicado por cinco las posibilidades de que sea culpable, pero continúa siendo más probable que sea inocente.

Supongamos, sin embargo, que obtenemos nuevas pruebas, se encuentra un pelo de caballo en el abrigo del sospechoso y la probabilidad de que el pelo del caballo pertenezca al que fue drogado es de 0,95 (hay solo cinco probabilidades entre cien de que el pelo pertenezca a un caballo diferente). En ese momento podríamos encadenar nuestras probabilidades bayesianas y cumplimentar una tabla nueva. Abajo del todo introducimos los valores que acabamos de calcular, 0,10 y 0,90 (a veces los estadísticos dicen que las posteriores de ayer son las previas de hoy). Si preferís pensar esas cifras como «una probabilidad entre diez» y «nueve probabilidades entre diez», podéis hacerlo y las colocáis como números enteros.

		Sospechoso culpable		
		SÍ	NO	
Coincidencia de las pruebas	SÍ	0,95	0,45	1,4
	NO	0,05	8,55	8,6
		1	9	10

Nuestros expertos forenses nos han informado de que la probabilidad de una coincidencia de la muestra de pelo es 0,95. Multiplicándolo por uno obtenemos la entrada de la celda superior izquierda y restándolo de uno obtenemos la entrada de la inferior izquierda. Si hay una probabilidad de 0,95 de que la muestra coincida con la del caballo drogado, eso significa que hay una probabilidad de 0,05 de que la muestra coincida con la de otro

animal (lo que absolvería al sospechoso), por lo tanto la celda superior derecha es el producto de 0,05 y el total marginal de 9 = 0,45. Al realizar nuestros cálculos observamos que

P(Culpable | Prueba) = 0,68	P(Prueba | Culpable) = 0,95
P(Inocente | Prueba) = 0,32	P(Prueba | Inocente) = 0,05

Las nuevas pruebas nos muestran que hay aproximadamente el doble de probabilidades de que el sospechoso sea culpable de que sea inocente, dadas las pruebas. Muchos abogados y jueces no saben organizar las pruebas de este modo, pero podéis comprobar lo útil que resulta. El problema de creer erróneamente que P(Culpable | Prueba) = P(Prueba | Culpable) está tan divulgado que ha sido denominado *falacia del fiscal*[198].

Si lo preferís, la aplicación del teorema de Bayes se puede realizar matemáticamente, en lugar de mediante la tabla de doble entrada, tal y como se muestra en el Apéndice.

---

195 K. Inman y N. Rudin (2002), «The origin of evidence», *Forensic Science International,* 126(1), 11-16.

K. Inman y N. Rudin (2000), *Principles and Practice of Criminalistics: the profession of forensic science,* Boca Raton, FL, CRC Press.

196 Fundamento esta sección en los comentarios de C. G. G. Aitken y F. Taroni (2004), *Statistics and the Evaluation of Evidence for Forensic Scientists,* 2.ª ed., Chicester, UK, John Wiley & Sons, pp. 1-2, y utilizo su contexto y sus términos.

197 C. G. G. Aitken y F. Taroni (2004), *op. cit.*

198 W. C. Thompson y E. L. Shumann (1987), «Interpretation of statistical evidence in criminal trials: the prosecutor's fallacy and the defense attorney's fallacy», *Law and Human Behavior,* 2(3), 167-187.

CUATRO ESTUDIOS DE CASO

La ciencia no nos ofrece certezas, solo probabilidades. No podemos estar seguros al cien por cien de que el sol saldrá mañana o de que el imán que elegimos atraerá al acero o de que nada es capaz de viajar a mayor velocidad que la luz. Creemos que todas esas cosas son muy probables, pero la ciencia únicamente nos ofrece las mejores conclusiones bayesianas disponibles, dado lo que conocemos hasta el presente.

El razonamiento bayesiano nos pide que consideremos las probabilidades a la luz de lo que conocemos sobre el mundo. Para afirmar ese compromiso resulta crucial el pensamiento crítico que describimos en esta guía de campo. El pensamiento crítico puede enseñarse y practicarse y es una habilidad que puede perfeccionarse. El estudio riguroso de casos particulares es uno de los enfoques habituales, ya que nos permite practicar en contextos nuevos lo que hemos aprendido, lo que los teóricos del aprendizaje denominan *transferencia lejana*. La transferencia lejana es el procedimiento más eficaz para consolidar el conocimiento.

Hay una variedad infinita de modos mediante los que se nos pueden colar razonamientos incorrectos e información errada. Nuestros cerebros no están construidos para la excelencia en este ámbito. El papel de la ciencia siempre ha sido tomar distancia y proceder a un razonamiento minucioso, sistemático. Los estudios de caso se presentan como narraciones, basados en sucesos reales parcial o totalmente y no cabe duda de que somos una especie a la que le encantan las narraciones. Recordamos las historias y el curioso modo con el que se engarzan con los conceptos fundamentales. Consideremos los siguientes como conjuntos de problemas que podemos explorar juntos.

Sombra, el perro maravilloso, tiene cáncer (¿o no?)

Adoptamos a nuestro perro Sombra, un cruce de pomerania y shetland, en un refugio para perros, cuando tenía dos años. Le pusimos ese nombre porque nos seguía de habitación en habitación por la casa durante el día, nunca se alejaba demasiado. Tal como ocurre con frecuencia con las mascotas, nuestros ritmos se sincronizaron, dormíamos y nos despertábamos a las mismas horas, teníamos hambre a las mismas horas y nos apetecía hacer ejercicio a las mismas horas. A menudo nos acompañaba en viajes de negocios a otras ciudades y se acostumbró a los aviones, trenes y automóviles.

Cuando Sombra tenía trece años comenzó a tener problemas de micción y una mañana encontramos sangre en su orina. Nuestro veterinario realizó un examen de ultrasonidos y descubrió un tumor en su vejiga. La única manera de saber si era canceroso era realizar dos procedimientos quirúrgicos que el oncólogo recomendaba: una cistoscopia, que supone introducir una cámara en miniatura por la uretra hasta la vejiga, y una biopsia, para obtener una muestra del tumor y analizarla al microscopio. Sin embargo, el veterinario nos advertía de los riesgos de utilizar anestesia en un perro de la edad de Sombra. Si el tumor resultaba ser canceroso, el oncólogo debería realizar una operación quirúrgica y comenzaría con la quimioterapia. Aun sin otras pruebas, los doctores estaban bastante seguros de que se trataba de un cáncer de vejiga, conocido como carcinoma de células transicionales (CCT). Tras un diagnóstico de ese tipo los perros sobreviven un promedio de seis meses.

Mi mujer y yo nos sentíamos completamente impotentes cuando mirábamos a los ojos de Sombra. No sabíamos si le dolía y, si tal era el caso, cuánto más tendría que soportar, fuera a causa del tratamiento o de la enfermedad. Nuestra responsabilidad era cuidarlo y eso suponía adoptar decisiones bajo presión emocional, lo que no implica deshacerse de la racionalidad. Es posible pensar críticamente bajo presión emocional, incluso cuando es vuestro perro.

Se trata de un escenario médico típico para las personas y para las mascotas: dos doctores, dos opiniones diferentes y muchas preguntas. ¿Cuáles son los riesgos de la cirugía? ¿Cuáles son los riesgos de la biopsia? ¿Cuánto tiempo podría sobrevivir Sombra si lo operaban y cuánto si no lo operaban?

En una biopsia se utiliza una aguja diminuta para retirar una muestra de tejido que se envía al patólogo, quien determina la probabilidad de que sea canceroso o no. (La patología, como la mayoría de las ciencias que hemos comentado, no trata de certezas, sino de posibilidades y probabilidades de que la muestra sea cancerosa, y, después, de que las partes no muestreadas del órgano también tengan cáncer; si buscáis certezas, la patología no es el lugar adecuado). Los pacientes y los dueños de mascotas casi nunca preguntan por los riesgos de la biopsia. En el caso de los humanos esa estadística se conoce bien, no así en veterinaria. La estimación que hacía nuestro veterinario era que había un 5% de probabilidad de una infección peligrosa y un 10% de probabilidad de que parte del material canceroso (si es que el tumor lo era) se extendiera por el abdomen al sacar la aguja, diseminando el cáncer. Un riesgo adicional es que las biopsias dejan un tejido cicatrizado que hace más difícil la cirugía posterior. La anestesia que requiere el procedimiento podía matar a Sombra. En pocas palabras: el procedimiento de diagnóstico podía hacer que empeorase.

Nuestro veterinario nos presentó seis opciones:

1. Biopsia a través de la pared abdominal con la esperanza de alcanzar un diagnóstico más definitivo.
2. Cateterismo diagnóstico (emplear un catéter para arañar una parte de la masa tumoral, lo que provoca la exfoliación de las células, que pueden ser analizadas).
3. Biopsia mediante el empleo de la propia cámara cistoscópica que de todos modos pretendían emplear para obtener una imagen mejor de la masa tumoral (a través de la uretra).
4. Cirugía mayor de inmediato a fin de ver directamente la masa tumoral y eliminarla en caso de que fuese posible. El problema es que la mayoría de los cánceres de vejiga recidivan a los doce meses, ya que los

cirujanos no consiguen retirar todas las células cancerosas y lo que queda sin retirar habitualmente crece rápidamente.

5. No hacer nada.
6. Sacrificar inmediatamente a Sombra, aceptando que lo más probable es que se trate de cáncer de vejiga y de todos modos no viviría mucho tiempo.

Le preguntamos cuáles eran las opciones de tratamiento si resultaba ser un cáncer y cuáles si no lo era. A menudo los pacientes se centran en el procedimiento a realizar de modo inmediato, sin averiguar cuáles pueden ser las siguientes etapas.

Si el tumor era canceroso, la principal inquietud era que pudiera crecer y bloquear uno de los canales que llevan la orina a la vejiga desde los riñones, o el que conduce la orina desde la vejiga hasta el césped o a la boca de incendios elegida a tal fin. Si ese bloqueo tenía lugar, Sombra sufriría un gran dolor y moriría en el plazo de un día. Además, podrían producirse bloqueos temporales como resultado de la inflamación. Dada la posición de la vejiga en relación al cuerpo y el ángulo de los ultrasonidos, resultaba difícil saber cuán cerca estaba el tumor de estos canales (el uréter y la uretra).

¿Qué hay de las seis opciones que presentamos más arriba? ¿Cómo decidir cuál de ellas elegir? Descartamos dos: sacrificar a Sombra y no hacer nada. Recordad que el oncólogo insistía en la cirugía ya que esa es su regla de oro, el protocolo habitual para tales casos. Le pedimos alguna estadística y nos dijo que lo indagaría y nos lo comentaría. Más tarde afirmó que existía un 20% de probabilidades de que la cirugía tuviera un mal desenlace y que Sombra muriera a resultas de la misma. En consecuencia, descartamos la cirugía mayor, ya que ni siquiera estábamos seguros de que el tumor fuera canceroso.

Solicitamos estadísticas de la esperanza de vida en los escenarios restantes. Lamentablemente los veterinarios no registran tales estadísticas y, en cualquier caso, las que se registran están sesgadas hacia una esperanza de vida breve, ya que muchos dueños de mascotas optan por la eutanasia. Es decir, muchos dueños deciden sacrificar a la mascota antes de

que la enfermedad avance, ya que les preocupa la calidad de vida del animal o su propia calidad de vida: los perros con CCT con frecuencia sufren incontinencia (ya nos habíamos dado cuenta de que Sombra nos dejaba pequeñas sorpresas en casa). Todavía no disponíamos de un diagnóstico definitivo, pero basándonos en las estadísticas a nuestro alcance, parecía que Sombra viviría tres meses *con o sin tratamiento*. Tres meses si no hacíamos nada, tres meses si le aplicábamos quimioterapia, tres meses si lo operábamos. ¿Cómo podían estar así las cosas? Según descubrimos, hacía diez años los veterinarios recomendaban la eutanasia en cuanto se diagnosticaba CCT. Y a la primera señal de incontinencia crónica, los dueños sacrificaban a sus perros. Por lo tanto, los dueños habitualmente terminaban con la vida de sus perros antes que el cáncer lo hiciera y eso provocaba que las estadísticas no fueran fiables.

Indagamos algo por nuestra cuenta, utilizando como términos de búsqueda «carcinoma de células transicionales» y «perro *o* canino». Encontramos que había un 30% de probabilidades de que Sombra mejorase simplemente ingiriendo un antiinflamatorio no esteroideo llamado Piroxicam. El propio Piroxicam presentaba efectos secundarios, entre los que se incluían malestares estomacales, vómitos, pérdida del apetito y problemas de riñón e hígado. Preguntamos al veterinario y estuvo de acuerdo en que tenía sentido comenzar con el Piroxicam, independientemente de cualquier otra cosa que decidiéramos hacer.

En la web de la Universidad de Purdue, uno de los centros más importantes de veterinaria clínica, pudimos obtener las siguientes estadísticas de supervivencia:

1. Mediana de supervivencia tras una cirugía mayor: 109 días
2. Mediana de supervivencia con quimioterapia: 130 días
3. Mediana de supervivencia con Piroxicam: 195 días

Sin embargo, el rango de tiempos de supervivencia de todos estos estudios variaba enormemente de perro a perro. Algunos de ellos morían al cabo de unos pocos días, otros vivían más de dos años.

Decidimos que la alternativa más racional era comenzar con Piroxicam, ya que sus efectos secundarios eran relativamente menores en comparación con los demás tratamientos, y obtener una cistoscopia para que el doctor pudiera ver el tumor más de cerca y la biopsia asociada, a fin de disponer de más datos sobre los que decidir. Sería necesario anestesiar levemente a Sombra, pero solo por un periodo breve y los doctores confiaban en que saldría bien del paso.

De hecho, dos semanas más tarde la cistoscopia mostró que el tumor se encontraba muy cerca de los uréteres y de las vías uretrales, de hecho estaba tan cerca que la cirugía no supondría una ayuda si el tumor resultaba ser canceroso, ya que buena parte del mismo quedaría donde estaba. El patólogo no pudo decidir si era canceroso o no, ya que no fue posible obtener una muestra de tejido lo suficientemente grande. Así que, después de todo, aún no teníamos un diagnóstico. Sin embargo, las estadísticas sugerían que si Sombra se encontraba en el 30% de los perros para quienes funcionaba el Piroxicam, ahí estaba su mejor esperanza de vida. No sería necesario someterlo a las incomodidades de la cirugía o la quimioterapia y podríamos disfrutar de su compañía en casa.

Tanto en mascotas como en humanos existen muchos casos en los que un tratamiento no mejora estadísticamente la esperanza de vida. Tomar estatina si no se pertenece a un grupo de alto riesgo o eliminar quirúrgicamente la próstata si no se padece un cáncer de desarrollo rápido son tratamientos que tienen un impacto nada desdeñable sobre la esperanza de vida. Suena contraintuitivo, pero es cierto: no todos los tratamientos suponen realmente una ayuda. Está claro que Sombra estaría mejor sin la cirugía (así se evitaba el 20% de probabilidades de que esta acabara con él) y en términos estadísticos la quimioterapia no le haría ganar tiempo.

Sombra respondió muy bien al tratamiento con Piroxicam y en tres días volvió a ser el que era: activo, de buen humor, feliz. Al cabo de una semana ya no tenía dificultades para orinar. Ocasionalmente observamos pequeñas cantidades de sangre en la orina, pero nos dijeron que eso era normal tras la biopsia. Más

adelante, 161 días después de tener las primeras sospechas (que nunca se confirmaron) de CCT, sus riñones empezaron a fallar. Hicimos que lo analizaran en una clínica oncológica especializada. Los doctores no tenían claro si el fallo orgánico estaba relacionado con la CCT o con otras causas. Prescribieron fármacos para abordar los trastornos habituales del riñón y realizaron docenas de pruebas sin llegar a averiguar qué sucedía. Sombra comenzó a sentirse peor y dejó de comer. Le pusimos un gotero con analgésico y dos días después, cuando le quitamos el gotero durante unos minutos para ver cómo iba, quedó claro que sufría dolores. Hablamos con los doctores que lo habían tratado al principio y con los actuales, describimos claramente su estado, su evolución y su situación actual. Todos estaban de acuerdo en que había llegado el momento de sacrificarlo. Disfrutamos de su compañía, y él de la nuestra, durante un mes más que un paciente promedio tratado con quimioterapia y durante el último mes fue posible evitar los hospitales, los catéteres, las vías intravenosas y los escalpelos.

Acudimos al hospital oncológico, el personal nos conocía bien porque habíamos llevado a Sombra todos los días entre sus pruebas y los tratamientos y dispusimos que lo sacrificaran. Sufría y pensamos que tal vez habíamos esperado dos o tres días de más. Fue doloroso ver cómo esa enorme personalidad se alejaba y desaparecía. Nos tranquilizó saber que habíamos tenido en cuenta todas las opciones para cuidar de él y que había disfrutado de una vida tan buena como pudimos proporcionarle el mayor tiempo posible. Quizá la emoción más difícil a que las personas se enfrentan cuando una enfermedad termina con una vida es arrepentirse de las decisiones que han adoptado. Nosotros pudimos despedirnos de Sombra sin arrepentirnos de nuestras decisiones. Permitimos que nos guiara nuestro pensamiento crítico, nuestro uso del razonamiento bayesiano.

¿Eran actores Neil Armstrong y Buzz Aldrin?

Quienes niegan la llegada a la Luna señalan ciertas inconsistencias y preguntas por responder. «Tendría que haberse producido una demora de más de dos segundos en las comunicaciones entre la Tierra y la Luna a causa de la distancia», «la calidad de las fotografías es inverosímilmente buena», «no aparecen estrellas en las fotos del cielo», «¿cómo puede haber ondeado la bandera de Estados Unidos, como si estuviera al viento, si en la Luna no hay atmósfera?». La guinda del pastel es un informe elaborado por un trabajador de la industria aeroespacial, Bill Kaysing, quien escribió que la probabilidad de un alunizaje con éxito era de 0,0017% (obsérvese la precisión de la estimación). Se han presentado muchas más objeciones. Una parte de lo que mantiene vivo al contraconocimiento es la mera cantidad de preguntas por responder que van surgiendo. Si queréis convencer a la gente de que algo no es verdad, aparentemente resulta muy eficaz acribillarla a preguntas y esperar que quede lo bastante impresionada, y abrumada, como para que no intente buscar explicaciones. Pero ni siquiera 1.000 preguntas por responder significan necesariamente que algo no sucedió, como sabe cualquier investigador. Las webs que se dedican a negar el alunizaje no citan las pruebas a favor, ni tampoco las refutaciones a sus afirmaciones.

En el caso del alunizaje cada una de las objeciones anteriores (y de muchas otras) se refuta muy fácilmente. Sí se produjo una demora de dos segundos en las comunicaciones, que puede apreciarse fácilmente en las grabaciones originales, pero algunos documentales y reportajes la eliminaron para hacer más impresionante la noticia. La calidad de las fotografías es alta[199] porque los astronautas utilizaron una cámara Hasselblad de alta resolución con película de alta resolución. No aparecen estrellas en el cielo lunar porque la mayoría de las imágenes fueron tomadas durante el día lunar (de lo contrario, no habría sido posible ver a los astronautas). Y la bandera no ondea: sabiendo que no había atmósfera, la NASA preparó una bandera con una vara de sujeción horizontal y las «ondas» son meras arrugas del tejido. Cómo no hay viento, estas no se mueven. Esta objeción se basa en fotos fijas en las que parece producirse un efecto de

ondulación, pero en la película se observa que la bandera no ondea, está estática.

¿Y qué hay del informe del trabajador de la industria aeroespacial sobre la extrema improbabilidad de un alunizaje? Para empezar, el «trabajador de la industria aeroespacial» no era un ingeniero ni un científico, sino un escritor con una licenciatura en inglés que trabajaba para Rocketdyne. La fuente de sus estimaciones parece ser un informe de Rocketdyne de los años cincuenta, cuando la tecnología espacial aún estaba en pañales. Aunque aún restan preguntas por responder (por ej., ¿porqué faltan algunos de los registros telemétricos originales?), las pruebas son completamente abrumadoras en el sentido de que el alunizaje fue real. No es una certeza, pero es muy muy probable. Si fuéramos a utilizar estimaciones de probabilidad espurias para afirmar que sucesos pasados no han tenido lugar, acabaríamos concluyendo que los seres humanos no existen realmente: se ha sostenido que las probabilidades de que se forme vida en la Tierra son de una entre muchos miles de millones[200]. Como sucede en muchos otros ejemplos de contraconocimiento, en este se utiliza el lenguaje de la ciencia, en este caso la probabilidad, a fin de degradar groseramente esa elegante lengua.

Estadísticas en escena (y en una caja)

David Blaine es un conocido mago e ilusionista. Afirma haber realizado hazañas inauditas de resistencia física (al menos una de ellas ha sido reconocida en el *Libro Guinness de los Récords*). Lo que interesa a un pensador crítico es: ¿realmente ha demostrado David Blaine una excepcional resistencia física o ha utilizado un truco inteligente? No cabe duda de que, como mago con experiencia, le resultaría fácil simular la resistencia física.

En una charla TED[201] que ha tenido más de 10 millones de visualizaciones[202], afirma haber sido capaz de contener la respiración bajo el agua durante más de diecisiete minutos y nos cuenta cómo se entrenó para lograrlo. También afirma haberse congelado en un bloque de hielo durante una semana, haber

ayunado durante cuarenta y cuatro días dentro de una caja de cristal y haber estado enterrado en un ataúd durante una semana. ¿Son ciertas estas afirmaciones? ¿Son siquiera verosímiles? ¿Hay alguna explicación alternativa?

En sus vídeos, Blaine actúa de modo sensato, no habla aceleradamente, no parece manipulador. Resulta creíble porque su discurso tiene la torpeza necesaria para hacernos difícil imaginar que está calculando lo que va a decir y cómo decirlo. Pero es necesario tener en mente que los magos profesionales habitualmente calculan y planifican todo lo que dicen. Cada uno de sus movimientos, cada vez que se rascan la cabeza de modo aparentemente espontáneo, ha sido ensayado una y otra vez. La ilusión que intentan crear, la hazaña de la magia, funciona porque el mago es un especialista en desviar nuestra atención y confundir nuestras impresiones sobre lo que es espontáneo y lo que no lo es.

¿Qué podemos hacer para aplicar el pensamiento crítico a sus sesiones de resistencia física?

Si apelamos a la jerarquía de fuentes de calidad, debemos considerar el hecho de que se trate de una charla TED y las charlas TED son sometidas a un control de verificación y su contenido es cuidadosamente corroborado. Pero ¿es siempre así? De hecho hay más de 5.000 eventos TED[203], pero únicamente son controlados con rigor dos de ellos: TED y TED Global. El vídeo de Blaine tiene su origen en una charla que impartió en TEDMED, una de las 4.998 conferencias organizadas por entusiastas y voluntarios que no son controladas directamente por la organización TED. Eso no significa que no sea verdad, sino que no podemos basarnos en la reputación y la autoridad de TED para establecer que sea verdad. Recordad cuando TMZ informó de la muerte de Michael Jackson, a veces dicen la verdad, incluso puede que lo hagan la mayor parte del tiempo, pero no podemos estar seguros.

Antes de considerar cómo contuvo la respiración bajo el agua, vamos a analizar con atención otro par de afirmaciones de Blaine. Para comenzar, la cadena de televisión Fox informó de que su demostración del bloque de hielo era un fraude[204]. Según indicó

la Fox, una trampilla bajo la habitación permitía trasladarse a un cuarto caliente y cómodo, mientras un doble tomaba su lugar en el bloque de hielo. ¿Cómo pudo colar ese engaño? Muchos magos practican repetidamente cómo conseguir que la audiencia acepte hechos que escapan a lo ordinario. Hay algunos indicios que llevan a pensar que las cosas no son lo que parecen. Por ejemplo, ¿lleva una máscara? (podríamos pensar que forma parte de la actuación o porque lo hace parecer tremebundo. La razón real podría ser porque facilita que nos engañe con un doble). ¿Para qué necesitan rociar sábanas con agua sobre el hielo cada cierto tiempo? (Blaine dice que es para impedir que el hielo se derrita; podría que ser que sirviera para cambiar de lugar con el doble durante los instantes en que no se ve a través del hielo). En cuanto al equipo de registro fisiológico que lleva adosado al cuerpo y que informa sobre la frecuencia cardíaca y la temperatura corporal, seguro que es real, ¿no? (¿y quién dice que el equipo está verdaderamente adosado a su cuerpo? Podría no ser así, y estar conectado a un ordenador).

Si Blaine mentía sobre el bloque de hielo, afirmando que se trataba de una hazaña de rendimiento físico, cuando en realidad se trataba de una prestidigitación, de un truco de magia, ¿no estará mintiendo sobre otras hazañas de rendimiento físico? Como artista con un gran número de seguidores querrá asegurarse de que sus actuaciones funcionan siempre. Emplear ilusionismo y trucos puede ser más fiable y seguro que llevar la resistencia hasta el límite. Pero incluso si se trata de trucos, quizá sea demasiado duro calificarlos de mentiras; al fin y al cabo, son parte del espectáculo, ¿no? Nadie cree que los magos se conecten realmente con las fuerzas de lo desconocido; sabemos que ensayan y practican constantemente y que utilizan la distracción de la atención. ¿A quién le importa? La mayoría de los magos prestigiosos, cuando se les pregunta, se sinceran y admiten que lo que hacen es ilusionismo, no demostraciones de magia negra. Por ejemplo, Glenn Falkenstein realizaba un número de lectura de mente considerado de los más impresionantes que se han llevado a cabo. Pero al final de cada actuación señalaba inmediatamente que no se trataba de una

verdadera lectura de mentes. ¿Por qué? Pues por su sentido de la ética[205]. El mundo está lleno de personas que creen cosas que no son ciertas y muchos creen cosas que son ridículas, afirmaba. Millones de personas tienen una comprensión limitada de las relaciones causa-efecto y malgastan su dinero y su energía en videntes, astrólogos, apuestas y terapias «alternativas» que carecen de eficacia probada. Ser sincero sobre cómo se lleva a cabo ese tipo de actuación es importante, decía, para que las personas no crean cosas que no son ciertas.

En otra de sus demostraciones Blaine afirma haberse clavado una aguja que atraviesa su mano. ¿Se trata de una ilusión o realmente lo hizo? No cabe duda de que en los vídeos parece real, pero claro, la magia es justamente eso. (Buscad en Youtube y encontraréis vídeos que muestran cómo hacerlo con un aparato especial). ¿Y qué hay de los cuarenta y cuatro días de ayuno en una caja de cristal? Hay incluso un artículo publicado en una revista con revisión por expertos[206] —el *New England Journal of Medicine*[207]—, y en cuanto a fuentes, es lo mejor que puede consultarse como fuente de información fiable. Sin embargo, si se analiza el trabajo con cuidado, se observa que los médicos que escribieron el artículo tan solo examinaron a Blaine tras el ayuno, no antes o durante el mismo, de modo que no pueden proporcionar una verificación independiente de que realmente ayunó. ¿Se abordó esa cuestión cuando el artículo fue revisado por expertos? El director actual de la revista buscó en los archivos[208], pero los registros habían sido destruidos, pues el artículo fue publicado una década antes de mi indagación. El autor principal del trabajo me contó por correo electrónico[209] que basándose en las hormonas que midió tras los hechos, estaba seguro de que había ayunado, pero también era posible que pasara bajo cuerda algo de comer, eso no podía asegurarlo. También llamó mi atención sobre un artículo de una colega suya, publicado en otra revista con revisión por expertos, en el que un médico *sí* que había supervisado a Blaine durante el ayuno[210] (el artículo no apareció en mis búsquedas en PubMed y Google Scholar porque a David Blaine no se le mencionaba por su

nombre). El siguiente pasaje del artículo, que apareció en *Nutrition,* resulta relevante:

> Inmediatamente antes de comenzar el ayuno, DB parece haber realizado ejercicios de desarrollo muscular congruentes con su índice de masa corporal, con los datos de su constitución física y con la circunferencia de la musculatura de los brazos, de los que se informa más abajo. La tarde del sábado, 6 de septiembre de 2003, DB ingresó en una caja transparente de metacrilato que medía 2,1 × 2,1 × 0,9 metros, suspendida en el aire durante los siguientes 44 días, cerca del Puente de la Torre, en Londres. Se puso a disposición de uno de los investigadores un registro de vídeo continuo y detallado (ME, en la oficina y en su residencia), quien pudo valorar el estado clínico y la actividad física de DB. DB, cuya edad era de 30 años, había consumido antes de los hechos una dieta que se estimó, aunque no pudo verificarse, había incrementado su peso en 6-7 kilos. También consumió compuestos multivitamínicos durante varios días antes del evento e interrumpió su consumo al entrar a la caja. Fue sintiéndose más débil y aletargado según avanzaba el proceso. A partir de la segunda semana experimentó algunos mareos y debilidad si intentaba levantarse con rapidez y ocasionales problemas temporales de visión, como si «fuera a desmayarse». También padeció dolores agudos y punzantes en los miembros y en el torso, malestar abdominal, náuseas y alguna arritmia cardiaca... Al quinto día sufrió una pequeña hemorragia nasal que se reprodujo más adelante, no hubo otras señales evidentes o síntomas de tendencia a sangrar. Tampoco se produjeron signos de edema al principio o al final del ayuno. Tampoco hubo signos clínicos de deficiencia de tiamina. DB, que al comienzo era un hombre de apariencia musculada, estaba visiblemente más delgado al salir de la caja. Su presión sanguínea medida casi inmediatamente antes del evento era de 140/90 mmHg echado y 130/80 mmHg en pie, y al final de 109/74 mmHg echado (89 pulsaciones/minuto) y de 109/65 de pie (119 pulsaciones/minuto).

A partir de este informe parece como si realmente hubiera ayunado. Un escéptico podría desdeñar sus informaciones sobre los dolores y las náuseas como algo impostado, pero es difícil amañar un ritmo cardíaco irregular y la pérdida de peso.

Sin embargo, eje de la charla TEDMED de Blaine es la contención de la respiración, televisada en el programa de Oprah

Winfrey. En ella Blaine utiliza mucho léxico científico y médico para apoyar la historia de que se trató de una demostración de resistencia física con fundamento científico y no un simple truco. Blaine describe la investigación que realizó:

«Me reuní con un neurocirujano experto y le pregunté durante cuánto tiempo... todo lo que esté por encima de seis minutos supone un riesgo serio de daño hipóxico cerebral... perflubron». Blaine menciona la respiración líquida; una cámara hipóxica a fin de desarrollar el cómputo de glóbulos rojos de la sangre; 02 puro. Eso le permitió aguantar quince minutos. Continúa comentando un régimen de entrenamiento mediante el que gradualmente consiguió alcanzar los diecisiete minutos. Coloca términos como «desviación de la sangre» e «isquemia». ¿Realmente llevó a cabo Blaine lo que cuenta? ¿La jerga médica que emplea tenía sentido, o se trata de farfulla pseudocientífica que invoca para impresionar, para hacernos *creer* que sabe de lo que está hablando?

Como siempre, vamos a empezar con una prueba de verosimilitud. Si alguna vez habéis intentado contener la respiración, probablemente lo habréis conseguido durante medio minuto, quizá incluso un minuto. Un poco de indagación nos revela que los pescadores profesionales de perlas habitualmente son capaces de contener la respiración durante siete minutos. El récord mundial de contención de la respiración hasta que llegó Blaine se situaba justo por debajo de diecisiete minutos. Si seguimos leyendo sobre el asunto descubrimos que existen dos tipos de competición de contención de la respiración: la contención de toda la vida, no respirar, como la que hacíamos con nuestro hermano en la piscina del barrio cuando éramos niños, y la contención *asistida,* en la que a los competidores se les permiten cosas como inhalar oxígeno cien por cien puro durante media hora antes de la competición. Eso empieza a sonar más verosímil, pero ¿hasta dónde es posible llegar con la contención asistida de la respiración? ¿Realmente se puede pasar de unos minutos hasta los diecisiete? Llegados a este punto nos interesaría saber qué pueden contarnos los expertos, los neumólogos (que tienen conocimientos sobre la capacidad

pulmonar y el reflejo respiratorio) y los neurólogos (que saben cuánto puede soportar un cerebro sin la entrada de oxígeno). Los dos neumólogos que consulté describieron un régimen de entrenamiento muy semejante al que describe Blaine en su vídeo; ambos pensaban que con esos «trucos» o medidas especiales resultaría posible aguantar la respiración durante diecisiete minutos. De hecho, el récord de Blaine fue superado en 2012 por Stig Severinsen[211], quien contuvo su respiración durante veinte minutos y diez segundos (desde luego, tras inhalar oxígeno puro) y quien un mes más tarde volvió a superar su propio récord, alcanzando veintidós minutos. El doctor David Eidelman, especialista en neumología y decano de la Escuela de Medicina de McGill, comentó: «Coincido en que resulta difícil de creer... no obstante, al inhalar primero oxígeno, ayunar y utilizar técnicas de yoga para disminuir la tasa metabólica mientras se contiene la respiración al sumergirse, parece que resulta posible. En consecuencia, pese a que mantengo cierto escepticismo, no creo que pueda demostrar que es imposible».

Charles Fuller, doctor y neumólogo en la Universidad de California en Davis, añade: «Me parece que hay pruebas suficientes para establecer que Blaine no miente, y que esa hazaña es fisiológicamente viable. Aunque haya que mantener la cautela de que Blaine es un mago y podría haber otros factores que contribuyan a que consiga contener la respiración durante diecisiete minutos, también parece haber suficientes pruebas fisiológicas de que sería posible realizar esa hazaña. En el mundillo de quienes contienen la respiración hay un pequeño grupo de personas que compiten por un récord oficialmente conocido como “apnea estática pre-oxigenada”. En ese evento la contención de la respiración está patrocinada por el Récord Mundial Guinness, ya que los buceadores deportivos lo consideran un truco. La duración de la contención de la respiración se mide tras hiperventilar (expulsar dióxido de carbono) durante treinta minutos, mientras se inhala oxígeno cien por cien puro. Es más, el evento habitualmente tiene lugar en una piscina de agua templada (lo que reduce la demanda metabólica de oxígeno), se contiene la respiración justo debajo de la

superficie, lo que induce el reflejo humano de inmersión (que contribuye a rebajar aún más la demanda metabólica de oxígeno). En otras palabras, todos estos “trucos” amplían la capacidad humana para contener conscientemente la respiración. Y, lo más importante, antes de Blaine, el récord se situaba justo por debajo de los diecisiete minutos (por un atleta que *no* era un mago) y existen otros individuos a quienes se ha registrado conteniendo la respiración durante periodos que superan los veinte minutos. Por lo tanto, hay pruebas suficientes de que esta hazaña podría haberse realizado tal y como se afirma».

Hasta ahora la historia de Blaine parece verosímil y su charla suena convincente. Pero ¿qué pasa con el daño cerebral? El propio Blaine lo cita como un problema. Seguro que habéis oído decir que si el cerebro se queda sin oxígeno durante tres minutos, puede producirse un daño irreversible y la muerte cerebral. Si uno no respira durante diecisiete minutos, ¿cómo se impide la muerte cerebral? Es una buena pregunta para un neurólogo.

Scott Grafton, doctor en Medicina: «El oxígeno no se mantiene en la sangre por sí solo. Imaginemos el agua y el aceite. Se escapa con rapidez del torrente sanguíneo, es necesario que se ligue a algo. La sangre contiene glóbulos rojos. Cada glóbulo rojo está cargado de moléculas de hemoglobina (Hgb). Esas moléculas de hemoglobina tienen la capacidad de ligar hasta cuatro moléculas de oxígeno. Cada vez que un glóbulo rojo circula por los pulmones, aumenta la cantidad de moléculas de oxígeno que se ligan a él. Cuanto mayor es la concentración de oxígeno en el aire, más moléculas de Hgb se vinculan al glóbulo. Así que si las cargamos, si respiramos un cien por cien de oxígeno durante treinta minutos, ligamos el oxígeno lo más cerca posible del cien por cien de saturación.

»Cada vez que un glóbulo rojo pasa por el cerebro, el oxígeno tiene una posibilidad de desligarse de la molécula, y penetrar a través de la membrana celular en el tejido cerebral, donde se liga a otras moléculas que lo utilizan en el metabolismo oxidativo. La probabilidad de que una cierta molécula de oxígeno se desligue de la hemoglobina y se difunda está en función de la diferencia

relativa de la concentración de oxígeno a cada lado de las membranas».

En otras palabras, cuanto más oxígeno necesita el cerebro, más probable resulta que lo extraiga de la hemoglobina. Al respirar oxígeno puro durante treinta minutos, se maximiza la cantidad de oxígeno que hay en el cerebro *y* en la sangre. A partir de ahí, una vez que comienza la competición, los niveles de oxígeno en el cerebro disminuyen, tal y como sucede normalmente, y el competidor extrae eficientemente el oxígeno que resta en la hemoglobina para oxigenar el cerebro.

Grafton continúa: «No todas las moléculas de hemoglobina se cargan de oxígeno al pasar por los pulmones y no todas se descargan cada vez que pasan por un órgano. Es necesario que pasen varias veces para descargarlas todas. Cuando afirmamos que la muerte cerebral tiene lugar rápidamente por la falta de oxígeno, suele ser en el contexto de falta de circulación (un ataque cardíaco), cuando el corazón ya no envía sangre al cerebro. Si el bombeo se detiene y no hay glóbulos rojos disponibles para distribuir el oxígeno, el tejido cerebral muere rápidamente. Cuando una persona se sumerge, hay una carrera entre el daño cerebral y el fallo del bombeo.

»Hay un truco que es clave: los músculos han de estar en reposo. Los músculos están cargados de mioglobina, que se liga al oxígeno cuatro veces más estrechamente que los glóbulos rojos. Si utilizamos los músculos se acelera la pérdida general de oxígeno. Hay que mantener baja la demanda muscular». Esta es la parte *estática* de la apnea estática que comentaba el Dr. Fuller.

En consecuencia, desde el punto de vista médico, las afirmaciones de David Blaine resultan verosímiles. Esto podría ser el final de la historia, si no fuera por lo siguiente. Un artículo publicado en el *Dallas Observer*[212] afirma que la contención de la respiración fue un truco y que Blaine, un ilusionista magistral, empleó un tubo de respiración muy bien oculto. La noticia no aparece en ningún otro medio de comunicación, lo que no significa que el *Observer* se equivoque, pero ¿por qué es este periódico el único que informa al respecto? Tal vez el que un

mago realice un truco y después afirme que no había tal no sea una noticia de interés.



El periodista John Tierney viajó a la isla Gran Caimán a fin de escribir sobre el entrenamiento de Blaine para un artículo del *New York Times*[213] y una semana más tarde escribió en su blog sobre la transmisión del asunto en el programa de *Oprah*[214]. Tierney concede gran importancia a la frecuencia cardíaca de Blaine, registrada en un monitor situado al lado del tanque de agua en el programa de Winfrey, pero igual que sucedía con la demostración del bloque de hielo, no hay pruebas de que el monitor realmente estuviera conectado a Blaine y podría haber estado allí más bien para dar juego al espectáculo, para que la audiencia creyera que las condiciones eran realmente complicadas (una práctica habitual entre los magos). Ni Tierney ni un médico implicado en el entrenamiento mencionan hasta qué punto supervisaron a Blaine durante sus ensayos de práctica en las islas Caimán; cabe que dieran por buena su palabra de no estar empleando ningún

aparato. Quizá el verdadero motivo de ese entrenamiento fuera que Blaine imaginó que si podía engañarlos a *ellos,* también podría engañar a la audiencia de la televisión. Tierney escribe[215]: «Yo estaba allí, acompañando a algunos buceadores expertos en apnea estática (contener la respiración mientras se mantiene inmóvil). El doctor Ralph Potkin, un neumólogo que estudia la contención de la respiración y que es médico del equipo de Estados Unidos de buceo libre, aplicó electrodos al cuerpo de Blaine durante la sesión y midió su respiración, su corazón y su sangre mientras Blaine mantuvo su cabeza sumergida en el agua durante dieciséis minutos.

»Siempre he sospechado de los embaucadores —hice un reportaje sobre James Randi[216] hace algún tiempo y estaba con él en Detroit cuando desenmascaró al evangelista Peter Popoff—, pero no vi razón alguna para poner en duda la hazaña de Blaine. Contuvo la respiración justo delante de mí, bajo un agua cristalina en la parte poco profunda de una piscina común del hotel, rodeado de expertos en contener la respiración que le observaron constantemente. Su nariz y su boca estaban claramente bajo el agua, pero solo unos centímetros, de modo que resultaban visibles todo el tiempo. Tendrías que explicarme cómo hubiera sido posible que escondiera un tubo sin que nadie se diera cuenta y sin que aparecieran burbujas. Los magos engañan a la gente distrayéndola con sus movimientos y su labia, pero el objetivo de la apnea estática es permanecer completamente inmóvil a fin de conservar el oxígeno, que es justo lo que hizo David. (La diferencia que esto supone es impresionante, los instructores que trabajaban con David hicieron una breve sesión de entrenamiento con mi fotógrafo y conmigo. Prerrespiramos aire en lugar de oxígeno, pero fue espectacular el tiempo que logramos aguantar; yo llegué a los 3 minutos y 41 segundos, y el fotógrafo aguantó todavía más)».

De modo que ahora el *Dallas Observer* decía que era un truco y un periodista del *New York Times* no creía que lo fuera. ¿Qué piensan al respecto los magos profesionales? Hablé con cuatro de ellos; el primero me dijo: «Tiene que ser un truco. En el gremio de la magia se sabe que en muchas de sus demostraciones

utiliza trucos con cámaras y una preparación muy elaborada. Le resultaría fácil disponer de un tubo de respiración que le permitiera aspirar oxígeno y exhalar dióxido de carbono sin dejar burbujas en el agua. Y si lo practica, no necesitaría hacerlo con mucha frecuencia, podría contener la respiración durante uno o dos minutos y respirar de vez en cuando por el tubo; además podría haber otro truco de cámara, ¡podría no estar ni siquiera en el agua! Una proyección o una pantalla verde podrían permitirle aparentarlo».

El segundo mago, que había trabajado con Blaine diez años antes, añadió: «Su héroe es Houdini, que se hizo famoso por llevar a cabo trucos peligrosos. Houdini en parte se labró su reputación haciendo el tipo de cosas que se hacía en los años veinte, sentarse en un mástil, etc. Algunas requerían resistencia y otras un truco fácil; algunas no eran tan difíciles como podría imaginarse, pero la mayoría de la gente no las intentaba nunca. No veo necesidad de que Blaine utilice un truco para el bloque de hielo, eso es fácil a causa del efecto iglú, realmente no hace tanto frío ahí dentro. *Parece* muy impresionante. Si estuviera dentro de un congelador sería distinto.

»Pero ¿diecisiete minutos conteniendo la respiración? Si puede hiperoxigenar su sangre eso ayudaría; me consta que entrena y que hace algunas cosas muy llamativas, pero estoy seguro de que lo de aguantar la respiración es, al menos parcialmente, un truco. Me parece que realmente contiene la respiración, pero no el cien por cien del tiempo. Es bastante sencillo de fingir. Probablemente usa un tubo para respirar u otro aparato.

»Observa que muchas de sus actuaciones las hace en la televisión y que hay partes editadas en los momentos clave. Tendemos a creer que se trata de información verídica y que lo estamos viendo todo, porque así es como nuestros cerebros elaboran la realidad. Pero se trata de un mago, veo que hay momentos editados y me pregunto qué es lo que sucede en el metraje que falta».

Un tercer mago añadió: «¿Para qué tomarse el trabajo de entrenar si, como ilusionista, puedes hacerlo ayudándote de

aparatos? Si utilizas aparatos generas una actuación más fiable, que puede repetirse fácilmente una y otra vez. Después, solo tienes que simular que estás dolorido, mareado, desorientado, como si hubieras arrastrado tu cuerpo más allá de cualquier límite razonable. Como artista no te interesa dejar nada al azar, hay demasiado en juego».

El hecho de que nadie informe de haber observado a David Blaine respirando mediante un tubo no es prueba de que no lo hiciera, ya que precisamente el *trabajo* de los ilusionistas consiste en realizar trucos basados en lo que vemos y en lo que creemos que vemos. La ilusión es incluso más potente cuando nos sucede a nosotros mismos. He visto al mago Glenn Falkenstein leer los números de serie de los billetes de dólar de mi cartera mientras estaba con los ojos vendados al otro lado de la habitación. El mago Tom Nixon puso en mi mano la carta del siete de diamantes y, sin embargo, pocos minutos después se había convertido en una carta completamente diferente, sin que me diera cuenta de que me tocaba o de que tocaba el naipe. Sé que en algún momento cambió la carta, pero incluso después de que me hiciera el truco cinco veces y habiendo observado cómo lo hacía con otras personas muchas más, sigo sin saber cómo lo consigue. Es parte del genio de los magos, forma parte del espectáculo. No creo ni por un instante que Falkenstein o Nixon tengan poderes ocultos, sé que se trata de un espectáculo y ellos lo consideran como tal.

El cuarto mago al que consulté fue James Randi, el escéptico profesional que John Tierney y yo mencionamos anteriormente, y que es capaz de reproducir presuntos fenómenos psíquicos mediante el empleo de hábiles ilusiones y trucos de magia. Esto es lo que me respondió por correo electrónico:

> Recuerdo cuando David Blaine comenzó a mostrar en televisión sus trucos, me puse en contacto con él para advertirle amistosamente de que, según mi opinión de ilusionista, se estaba exponiendo a sufrir daños físicos. Intercambiamos una correspondencia amistosa, hasta que fui abruptamente informado de que su nueva agencia de representación había cambiado su dirección de correo electrónico y le había indicado que dejara de mantener

correspondencia conmigo. No dudé en aceptar esa decisión, al tiempo que esperaba que el señor Blaine tuviera en cuenta mis bienintencionadas sugerencias.

No he tenido contacto con David Blaine desde entonces. Me alarmó conocer las imprudentes afirmaciones que realizó en la charla TED y he respetado el, para mí, sesgo poco prudente que su agencia ha decidido otorgar a sus declaraciones, pero he respetado su privacidad.

Permitió que su representante pusiera fin a su contacto conmigo, tal vez porque procuré que se mantuviera en el terreno de la honestidad. Nunca se es demasiado honesto.

Los hechos que hemos podido corroborar sugieren que contener la respiración durante diecisiete minutos es bastante verosímil, si bien eso no garantiza que Blaine no utilizase un tubo para respirar. Que creamos o no que Blaine ha realizado una actuación legítima queda al criterio de cada uno. Igual que sucede con cualquier mago, no podemos estar seguros de lo que es verdad y de lo que no lo es, ese es el mundo de ambigüedad que los magos se esfuerzan por crear y que constituye su profesión. Cuando pensamos críticamente, buscamos la explicación más sencilla, pero en algunos casos, como sucede en este, resulta difícil o imposible elegir entre las posibles explicaciones o imaginar cuál es la más sencilla. ¿Tiene alguna importancia? Pues la verdad es que sí. Como dijo Falkenstein, las personas que tienen una comprensión limitada de causas y efectos o no tienen un concepto claro del azar y la probabilidad son fácilmente engañadas mediante afirmaciones como las anteriores, lo que les conduce a aceptar otras con facilidad. Por no mencionar a los muchos aficionados que pueden intentar repetir esos espectáculos, pese a la omnipresente advertencia de «no intente hacerlo por su cuenta». Las personas con poca formación son blancos fáciles. La diferencia entre llevarlo a cabo con entrenamiento y realizarlo gracias a un truco es la diferencia entre ser engañado o no ser engañado.

Las estadísticas en el Universo

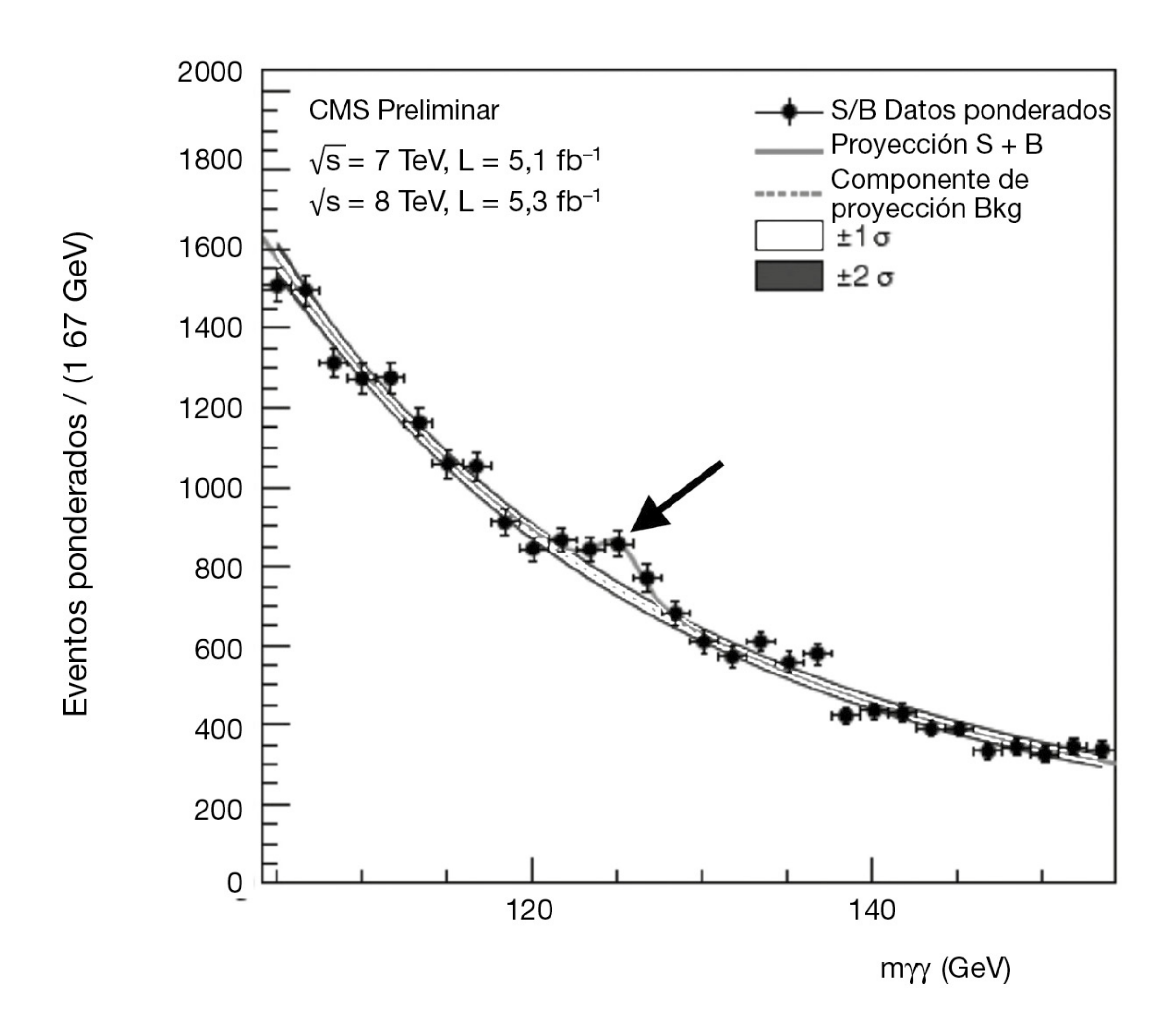
Cuando escuchamos términos como hidrógeno, oxígeno, boro, estaño y oro, ¿en qué pensamos? Se trata de elementos químicos de la tabla periódica, sobre los que habitualmente nos instruyen en la enseñanza secundaria. Los científicos los denominaron elementos porque se creía que eran unidades indivisibles fundamentales de la materia (del latín *elementum,* la materia en su forma más básica). El científico ruso Dmitri Mendeléyev identificó un patrón de propiedades en los elementos y los organizó en una tabla que permitiera visualizar más fácilmente esas propiedades. Al hacerlo, observó espacios de elementos que aún no habían sido descubiertos. Finalmente se han descubierto en la naturaleza o se han sintetizado en el laboratorio todos los elementos entre el 1 y el 118[217], lo que corrobora la teoría que subyace a la disposición de la tabla.

Más tarde los científicos descubrieron que en realidad los elementos químicos no eran indivisibles; están formados por lo que llamaban átomos, del griego *atomos,* que significa «indivisible». Pero también se equivocaban respecto a esa indivisibilidad. Más adelante se descubrió que los átomos están formados de partículas subatómicas: protones, neutrones y electrones. Inicialmente también se creía que estos eran indivisibles, pero... lo habéis adivinado, se descubrió que era incorrecto. El denominado modelo estándar de la física de partículas se formuló en los años cincuenta y sesenta y propuso que los electrones son indivisibles, pero que los protones y los neutrones se componen de partículas subatómicas más pequeñas. Cuando se descubrieron los quarks en los años setenta, el modelo fue confirmado. Para complicar aún más los términos, los protones, los neutrones y los electrones son un tipo de *fermión* y los fotones son un tipo de *bosón*. Estas distintas categorías son necesarias porque los dos tipos de partículas se rigen por leyes diferentes. Los fermiones y los bosones reciben el nombre de *partículas elementales* porque se cree que son realmente indivisibles (pero el tiempo lo dirá).

Según el Modelo Estándar hay diecisiete tipos diferentes de partículas elementales, doce tipos de fermiones y cinco tipos de bosones. El bosón de Higgs, que recibió mucha atención en la

prensa en 2012 y 2013, ha sido la última pieza del Modelo Estándar que se ha comprobado, las otras dieciséis ya se habían descubierto. Si existe, el bosón de Higgs ayudaría a explicar cómo obtiene masa la materia y cubriría un vacío crucial en la teoría que utilizamos para explicar la naturaleza del universo, un vacío en la teoría que ha estado presente durante más de cincuenta años.

¿Cómo sabemos si lo hemos encontrado? Cuando las partículas colisionan a gran velocidad... bah, olvidadlo. Vamos a dejar que nos lo explique un físico. Aquí tenemos al profesor Harrison Prosper[218] describiendo esta trama y la pequeña «irregularidad» señalada por la flecha que corresponde a 125 gigaelectronvoltios (GeV) en el eje horizontal:



El gráfico muestra «un espectro que surge de las colisiones protón-protón que dan lugar a la creación de un par de fotones

(gammas en la jerga de las altas energías)», dice Prosper. «El Modelo Estándar predice que el bosón de Higgs debería descomponerse (es decir, romperse) en un par de fotones. (También se ha predicho que el Higgs puede descomponerse de otros modos, como en un par de bosones Z). El bulto en la trama cerca de 125 GeV es una prueba de la existencia de alguna partícula de masa definida que se descompone en un par de fotones. Ese ‘algo’, hasta donde hemos podido averiguar, es probablemente el bosón de Higgs».

No todos los físicos coinciden en que esos experimentos sean concluyentes. Louis Lyons[219] afirma: «El Higgs... puede descomponerse en seis conjuntos distintos de partículas y estas ratios son definidas por el Modelo Estándar. Medimos esas ratios, pero con los datos actuales no hay ninguna seguridad. Son congruentes con las predicciones del ME, pero con más datos podría resultar mucho más convincente. De ahí la prudencia con la que puede afirmarse que hemos descubierto el Higgs del ME».

En otras palabras, los experimentos son tan caros y difíciles de realizar que los físicos prefieren evitar una falsa alarma; ya se han equivocado antes. Aunque los responsables del CERN anunciaron en 2012 que lo habían encontrado[220], muchos físicos piensan que el tamaño de la muestra era demasiado reducido. Hay tanto en juego que los físicos se han autoimpuesto un nivel de prueba, un umbral estadístico, que es mucho más estricto que el 1 sobre 20 habitual en otros ámbitos: 1 sobre 3,5 millones. ¿Por qué un requisito de prueba tan radical? Prosper afirma[221]: «Dado que la búsqueda del Higgs ha durado cuarenta y cinco años, que han participado miles de científicos e ingenieros, se han invertido miles de millones de dólares... por no mencionar los divorcios, las horas de sueño perdidas, las decenas de miles de comidas de avión de mala calidad, etc., queremos estar tan seguros como sea humanamente posible de que estamos en lo cierto».

El físico Mads Toudal Frandsen[222] añade: «Los datos del CERN en general se consideran prueba de que la partícula es la partícula de Higgs. Es cierto que la partícula de Higgs puede

explicar los datos, pero puede haber otras explicaciones; también podrían obtenerse esos datos a partir de otras partículas. Los datos actuales no son lo suficientemente precisos para determinar exactamente de qué partícula se trata. Podría tratarse de otras partículas conocidas». Recordad los comentarios previos de *La mentira como arma* sobre las explicaciones alternativas. Los físicos están alerta al respecto.

Si la trama muestra pruebas de un tipo diferente de partícula, alguna que *no* sea el de Higgs, eso podría suponer un cambio sustancial en nuestra concepción sobre cómo fue creado el universo. Y si existe, algunos físicos, como Stephen Hawking, temen que podría suponer el fin del universo tal como lo conocemos. El temor es que una fluctuación cuántica podría generar una burbuja de vacío que se expandiera rápida y continuamente hasta aniquilar el universo. Y si opináis que los físicos carecen de sentido del humor, Joseph Lykken[223], físico y director del Laboratorio del Acelerador Nacional Fermi en Illinois, observó que eso no sucederá hasta dentro de mucho, mucho tiempo, 10100 años a partir de ahora, «por lo tanto, es mejor que no vendáis vuestras casas y sigáis pagando vuestros impuestos».

No todo el mundo está contento con el descubrimiento y no porque sea una señal del fin del mundo, sino porque el hecho de que la ciencia descubra algo que la teoría estándar predice no abre las puertas a nuevas investigaciones. Un resultado anómalo e inexplicado es mucho más interesante para los científicos, ya que significa que su modelo y su enfoque en el mejor de los casos resultaban incompletos y, en el peor, estaban completamente equivocados, lo que supone una gran oportunidad para generar conocimientos nuevos. En una de las muchas intersecciones entre arte y ciencia, el director de orquesta Benjamin Zander comenta que cuando un músico comete un error, en lugar de maldecir o decir «ups» o «lo siento», debiera decir «¡vaya, *esto* es interesante!». Es interesante porque supone una oportunidad para aprender. Puede que el descubrimiento del bosón de Higgs dé respuesta a todas las preguntas que teníamos. O, como afirma el escritor de *Wired,* Signe Brewster[224],

«podría conducirnos a un principio subyacente que ha pasado desapercibido a los científicos hasta ahora. La meta final es tocar una cuerda que cuando se pulse haga sonar una campana de alarma que llame la atención de los físicos hacia algo nuevo». Como se dice que comentó Einstein, si sabes qué resultado va a dar algo, entonces no es ciencia, es ingeniería.

Los científicos aprenden a lo largo de la vida, son curiosos, están deseosos de afrontar el próximo desafío. Hay quienes temen que el descubrimiento del Higgs pueda explicar tanto que termine con el misterio. Otros están tan inmersos en las maravillas y complejidad de la vida y del universo que confían en que jamás lograrán explicarlo todo. Me cuento entre los segundos.

Según escribo estas líneas, han aparecido pruebas promisorias en el CERN sobre una partícula nueva que podría ser un gravitón, o una versión más pesada del bosón de Higgs. Pero la explicación más probable de estos picos en el flujo de datos es que se trate de una coincidencia: los resultados presentan una probabilidad de 1 sobre 93 de ser casualidad, mucho mayor que la probabilidad de 1 sobre 3,5 millones que se utiliza para el Higgs. Sin embargo, hay consideraciones cualitativas. «Lo bonito es que no se trata de una señal especialmente llamativa en un canal bastante claro» —declaró el físico Nima Arkani-Hamed al *New York Times*[225]—. Por lo tanto, aunque no vamos a meter el champán en la nevera, es interesante». Nadie sabe todavía de qué se trata y eso es justo lo que encanta a Lykken y a muchos otros que disfrutan con la emoción de la cacería.

La ciencia, la historia y las noticias están repletas de cosas que conocemos o creemos que conocemos, hasta que se descubre que nos equivocábamos. Un componente esencial del pensamiento crítico es saber que no sabemos. Un principio rector, por lo tanto, es darse cuenta de que sabemos lo que sabemos, hasta que nos damos cuenta de que no es así. El propósito de *La mentira como arma* es ayudar a pensar con atención, proporcionaros más confianza tanto sobre aquello que creéis conocer, como sobre aquello que pensáis que desconocéis y,

espero, a ser capaces de establecer la diferencia entre ambas cosas.

---

199 Hasselblad.com. https://www.hq.nasa.gov/alsj/a11/a11-hass.html; http://www.wired.com/2013/07/apollo-hasselblad/.

200 Entre la estimaciones se incluye 1 × 10390. http://evolutionfaq.com/articles/probability-life. Véase también D. Dreamer (30 de abril de 2009), «Calculating the odds that life could begin by chance», *Science 2.0,* http://www.science20.com/stars_planets_life/calculating_odds_life_could_begin_chance.

201 *N. del T.:* Las charlas TED (Tecnología, Entretenimiento y Diseño) se imparten por una organización sin ánimo de lucro, son mundialmente conocidas por la calidad y prestigio de sus ponentes (personajes como Bill Gates, Bill Clinton, premios Nobel, etc.) y abarcan temas muy amplios, como ciencias, arte y diseño, política, educación, cultura, negocios, asuntos globales, tecnología, desarrollo y entretenimiento.

202 https://www.ted.com/talks/david_blaine_how_i_held_my_breath_for__17min?language=en.

203 Bruno Guissani, Conservador de TEDGlobal Conference, comunicación personal, 28 de septiembre de 2015.

204 https://www.youtube.com/watch?v=U6Em2OhvEJY.

205 Glenn Falkenstein, comunicación personal, 25 de octubre de 2007.

206 M. Korbonits, D. Blaine, M. Elia y J. Powell-Tuck (2005), «Refeeding David Blaine. Studies after a 44-day fast», *New England Journal of Medicine,* 353(21), 2306-2307.

207 *N. del T.:* Una de las revistas de medicina más prestigiosas.

208 J. Drazen, Doctor en Medicina, comunicación por correo electrónico, 20 de diciembre de 2015.

209 M. Korbonits, Doctor en Medicina, comunicación por correo electrónico, 25 de diciembre de 2015.

210 J. M. Jackson, *et al.* (2006), «Macro- and micronutrient losses and nutritional status resulting from 44 days of total fasting in a non-obese man», *Nutrition,* 22(9), 889-897.

211 R. Grenoble (16 de noviembre de 2012), «Breathholding world record: Stig Severinsen stays under water for 22 minutes (video)», *Huffington Post,* http://www.huffingtonpost.com/2012/11/16/breath-world-record-stig-severinsen_n_7083351.html.

212 E. Liner (13 de enero de 2012), «Want to know how David Blaine does that stuff? (Don't hold your breath)», http://www.dallasobserver.com/arts/want-to-know-how-david-blaine-does-that-stuff-dont-hold-your-breath-7083351.

213 J. Tierney (22 de abril de 2008), «This time, he'll be left breathless», *New York Times,* p. F1.

214 J. Tierney (2008), «David Blaine sets breath-holding record», http://tierneylab.blogs.nytimes.com/2008/04/30/david-blaine-sets-breath-holding-record.

215 John Tierney, correspondencia por correo electrónico, 13 y 18 de enero de 2016.

216 *N. del T.:* Conocido ilusionista estadounidense, famoso por su escepticismo y su competencia a la hora de desenmascarar embaucadores.

217 D. Netburn (4 de enero de 2016), «It's offcial: four super-heavy elements to be added to the periodic table», http://www.latimes.com/science/sciencenow/la-sci-sn-new-elements-2016014-story.html.

218 H. B. Prosper (11 de julio de 2012), «International Society for Bayesian Analysis», http://bayesian.org/forums/news/3648.

219 L. Lyons (11 de julio de 2012), http://bayesian.org/forums/news/3648.

220 En artículos sobre el bosón de Higgs es posible encontrar referencias a un nivel de prueba de 5 desviaciones típicas. Las cinco desviaciones típicas hacen referencia al nivel de probabilidad que los científicos acuerdan antes de realizar un experimento —la probabilidad de que se realice una interpretación equivocada ha de tener un intervalo de confianza dentro de 5 desviaciones típicas (5-sigma) o 0,00000005 (recuérdese que previamente hablamos de intervalos de confianza del 95 y del 99%— en este caso es un intervalo de confianza del 99,99995%). Véase http://blogs.scientificamerican.com/observations/five-sigmawhats-that/.

221 H. B. Prosper (10 de julio de 2012), http://bayesian.org/forums/news/3648.

222 7 de noviembre de 2014, «Maybe it wasn't the Higgs particle after all. Phys.org», http://phys.org/news/2014-11-wasnt-higgs-particle.html.

223 http://phys.org/news/2014-11-wasnt-higgs-particle.html.

224 http://www.wired.com/2015/11/physicists-are-desperate-to-be-wrong-about-the-higgs-boson/.

225 D. Overbye (16 de diciembre de 2015), «Physicists in Europe find tantalizing hints of a mysterious new particle», *New York Times,* p. A16.

CONCLUSIÓN

DESCUBRIR POR SÍ MISMO

En *1984,* de George Orwell, el Ministerio de la Verdad era la agencia oficial de propaganda del país, responsable de falsificar los registros históricos y otros documentos de acuerdo con los propósitos de la administración. El Ministerio también promovía el contraconocimiento cuando convenía a sus intenciones, como 2 + 2 = 5.

1984 se publicó en 1949, medio siglo antes de que Internet se convirtiera en nuestra fuente de información de hecho. Hoy día, al igual que en *1984,* es posible modificar los sitios web de manera que el cambio pase inadvertido para el promedio de la gente; cada dato de la información puede reescribirse o (como en el caso de Paul McCartney y Dick Clark) mantenerse fuera de su alcance. Hoy día puede resultar muy difícil a quien navega por la red saber si una web está mostrando conocimiento real o contraconocimiento. Lamentablemente, a menudo los sitios en donde se afirma que cuentan la verdad son los que no la cuentan. En muchos casos la palabra «verdad» ha sido secuestrada por personas que divulgan contraconocimiento o puntos de vista marginales que son contrarios a lo que convencionalmente aceptamos como verdadero. Incluso los nombres de esos sitios pueden ser engañosos.

¿Podemos confiar en los expertos? Depende. La especialización tiende a ser estrecha. Un economista situado en los niveles más altos del gobierno puede carecer de conocimientos concretos sobre qué programas sociales resultarían eficaces para combatir la delincuencia. Y a veces los expertos son captados por determinados intereses y, sin duda, comenten errores.

Un sesgo anticiencia ha penetrado en el discurso público en la web. Muchos asuntos que debieran ser problemas científicos o técnicos[226], como dónde construir una central eléctrica y cuánto podría costar, son problemas políticos. Cuando tal cosa sucede, el proceso de toma de decisiones se pervierte y los hechos que realmente importan dejan de ser los que se toman en cuenta. O afirmamos que deseamos tratar una enfermedad humana incurable, pero nos burlamos cuando se da el primer paso y se invierten decenas de millones de dólares en el estudio de los áfidos. Lo cierto es que la ciencia progresa obteniendo conocimientos sobre la fisiología celular básica. Si se considera desde una perspectiva equivocada, la investigación puede parecer trivial; desde una perspectiva correcta puede contemplarse su verdadero potencial para resultar transformadora. El dinero que se invierte en los ensayos clínicos con humanos puede acabar permitiendo que se traten los síntomas de unos cuantos cientos de miles de personas. Ese mismo dinero invertido en investigación científica básica tiene el potencial de curar *docenas* de enfermedades de *millones* de personas, porque aborda los mecanismos comunes a muchos tipos diferentes de bacterias y de virus. El método científico es el terreno en el que surge el mejor pensamiento crítico.

Además del sesgo anticiencia, en Internet también se produce un sesgo antiescéptico. Muchas personas piensan «si lo he encontrado en línea, debe ser verdad». Al no existir ninguna autoridad responsable de supervisar y regular los sitios web y los demás contenidos que se encuentran en línea, la responsabilidad de verificar las afirmaciones recae sobre cada uno de nosotros. Afortunadamente existen algunas webs para ayudarnos. Snopes.com y otros sitios semejantes se dedican a desvelar leyendas urbanas y afirmaciones falsas. Empresas como Consumer Reports dirigen laboratorios independientes que proporcionan evaluaciones ecuánimes de diferentes productos, sin dejarse influir por lo que declaran los fabricantes. Consumer Reports ha existido durante décadas, pero hay buenos motivos para suponer que otras iniciativas de fomento del pensamiento crítico van a surgir en el siglo XXI. Esperemos que así sea. De

todos modos, por mucha ayuda que nos puedan proporcionar distintos medios, cada uno de nosotros tendrá que aplicar su propio criterio.

La promesa de Internet es que sea una gran fuerza democratizadora, que permita a todos expresar sus opiniones y ofrezca a todos acceso inmediato a la información mundial. Combinando ambas, como hacen Internet y las redes sociales, dispondréis de un mundo virtual en el que conviven juntas información y desinformación, mirándonos como gemelos idénticos: uno habrá de ayudarnos, el otro nos hará daño. Saber cuál elegir recae sobre nosotros y exige una consideración cuidadosa y algo de lo que muchos carecemos: tiempo. El pensamiento crítico no es algo que se realice una vez y se deje de lado. Es un proceso activo y regular. Requiere que todos nosotros pensemos como bayesianos y actualicemos nuestros conocimientos según incorporamos más información.

El tiempo que pasamos evaluando afirmaciones no sólo es un tiempo bien empleado, sino que debería considerarse parte de un trato implícito que todos hemos hecho. La recopilación y valoración de la información que solía demorarse desde horas a semanas, hoy nos cuesta unos segundos. Hemos ahorrado una cantidad incalculable de horas de acceso a bibliotecas, a archivos remotos, a rastrear libros voluminosos en busca de un párrafo que responda nuestras preguntas. El trato implícito que todos debemos hacer explícito es que utilizaremos *una parte* del tiempo que hemos ahorrado en la adquisición de información realizando una verificación adecuada de la información. Igual que resulta difícil fiarnos de alguien que nos ha engañado, es difícil confiar en nuestros conocimientos si resultan ser contraconocimiento. El hecho es que ahora mismo el contraconocimiento florece en Facebook, en Twitter y en los blogs... en todas las plataformas semiorganizadas.

Nos irá mucho mejor si conocemos una cantidad moderada de cosas con certidumbre que si conocemos muchas que son falsas. El contraconocimiento y la información errónea pueden resultar caros en términos de vidas y de felicidad y en términos del tiempo que hay que emplear para desmontar cosas que no funcionan

como esperábamos que funcionaran. El conocimiento verdadero simplifica nuestras vidas, nos ayuda a realizar elecciones que incrementan nuestra felicidad y nos ahorran tiempo. Seguir los pasos de *La mentira como arma* para evaluar la miríada de afirmaciones que nos encontramos es un modo de mantenerse dos pasos por delante de los millones de mentiras que hay en la web y por delante de los mentirosos y simples incompetentes que las perpetran.

---

226 Estas ideas y su formulación provienen de D. Frum (2015), «Charla impartida en Colleges Ontario Higher Education Summit», 16 de noviembre de 2015, Toronto, ON.

APÉNDICE

APLICACIÓN DEL TEOREMA DE BAYES

El teorema de Bayes puede expresarse como sigue:

$$P(A \mid B) = \frac{(P(B \mid A) \times P(A))}{(P(B))}$$

Para el problema presente vamos a emplear una notación según la cual G se refiere a la probabilidad previa de que el sospechoso sea culpable (antes de que conozcamos el informe del laboratorio) y E se refiere a la prueba de coincidencia de sangre. Queremos saber P(G | E). Sustituyendo en la fórmula de arriba, ponemos G en lugar de A y E en lugar de B y obtenemos:

$$P(G \mid E) = \frac{(P(E \mid G) \times P(G))}{(P(E))}$$

Para calcular el teorema de Bayes[227] y resolver P(G | E) puede ser útil emplear una tabla. Los valores son los mismos que usamos en la tabla de doble entrada de la página 229.

CÁLCULO DEL TEOREMA DE BAYES

Hipótesis (H) (1)	*Probabilidad previa P(G) (2)*	*Probabilidad prueba P(E | G) (3)*	*Producto (4)=(2)(3)*	*Probabilidad posterior P(G | E) (6)=(4)/Suma*
Culpable	0,02	0,85	0,017	0,104
Inocente	0,98	0,15	0,147	0,896
		Suma =	0,164 = P(D)	

Por lo tanto, redondeando, P(Culpable | Prueba) = 0,10
P(Inocente | Prueba) = 0,90

---

227 G. R. Iversen (1984), *Bayesian Statistical Inference,* Quantitative Applications in the Social Sciences, vol. 43, Thousand Oaks, CA, Sage.

GLOSARIO

Esta lista de definiciones no es exhaustiva, más bien se trata de una selección personal realizada a partir de la experiencia de escribir este libro. Sin duda podéis aplicar vuestro propio criterio independiente en este caso e identificar algunas definiciones que pueden ser mejoradas.

Abducción. Tipo de razonamiento popularizado por Sherlock Holmes mediante el que se utilizan intuiciones inteligentes para construir una teoría que explique los hechos observados.

Acumulativo, gráfico. Un gráfico en el que la cantidad que se mide, por ejemplo, las ventas o la afiliación a un partido político, es representada por el total actual, en lugar de por la cantidad de observaciones nuevas realizadas durante un periodo temporal. Lo ilustramos mediante las ventas acumuladas del iPhone (p. 67).

Afirmación del antecedente. Lo mismo que *modus ponens* (véase).

Agregar. Combinar observaciones o puntuaciones de dos o más grupos en uno único. Si los grupos son semejantes en una dimensión importante, si son homogéneos, habitualmente es lo correcto. Si no lo son puede provocar distorsiones de los datos.

Bimodal, distribución. Un conjunto de observaciones en el que dos valores tienen lugar más frecuentemente que otros. Una gráfica de sus frecuencias frente a sus valores muestra dos picos, o jorobas, en la distribución.

Condicional, probabilidad. La probabilidad de que un suceso ocurra *dado* que otro tenga lugar o haya ocurrido. Por ejemplo, la probabilidad de que llueva hoy dado que ayer

llovió. La palabra «dado» se representa mediante una línea vertical como esta: |.

Contexto. Modo mediante el que se informa de una estadística, por ejemplo, el contexto que se proporciona, o el grupo de comparación o agregación que se utiliza, pueden influir sobre cómo interpretamos una estadística. Comparar la cantidad total de accidentes aéreos que tuvieron lugar en 2016 con los que sucedieron en 1936 puede confundir, ya que hubo muchísimos más vuelos en 2016 que en 1936. Diversas medidas ajustadas, como la cantidad de accidentes por 100.000 vuelos o la cantidad de accidentes por cada 100.000 millas de vuelos realizados, proporcionan un resumen más exacto. Debemos esforzarnos por encontrar el contexto verdadero de cada estadística, es decir, el más adecuado e informativo. Calcular las proporciones en lugar de las cifras directas a menudo ayuda a establecer el contexto adecuado.

Contrapositivo. Un tipo de deducción válida que tiene la siguiente forma:

Si A, entonces B
No B
Por tanto, no A

Conversión, error de. Una forma no válida de razonamiento deductivo del siguiente tipo:

Si A, entonces B
B
Por tanto, A

Correlación. Medida estadística del grado en que dos variables se relacionan entre sí, su valor abarca desde –1 a 1. Una correlación perfecta (correlación = 1) tiene lugar cuando una variable cambia completamente con otra. Se produce una correlación negativa perfecta cuando una variable cambia de modo absolutamente contrario a otra

(correlación = –1). Se produce una correlación de 0 cuando dos variables carecen por completo de relación.

Una correlación indica únicamente que dos (o más) variables están vinculadas, no que una sea la causa de la otra. La correlación no supone causación.

La correlación también resulta útil porque proporciona una estimación de cuánta variabilidad de las observaciones se debe a las dos variables bajo seguimiento. Por ejemplo, la correlación de 0,78 entre estatura y peso indica que el 78% de las diferencias de peso entre los individuos están vinculadas a su estatura. El estadístico no nos indica a qué debe atribuirse el 22% restante de la variabilidad, para saber eso serían necesarios más experimentos, pero cabe imaginar que factores como la dieta, la genética, el ejercicio físico, etc., sean parte de ese 22%.

Cum hoc, ergo propter hoc (con esto, por tanto, debido a esto). Falacia lógica que surge de creer que simplemente porque dos sucesos ocurren juntos, uno ha de ser la causa del otro. La correlación no supone causación.

Deducción. Forma de razonamiento en la que se avanza desde la información general hasta una predicción específica.

Eje de y doble. Una técnica gráfica para agregar dos conjuntos de observaciones sobre el mismo gráfico, en el que se representan los valores de cada conjunto en dos ejes diferentes (habitualmente con escalas diferentes). Únicamente resulta apropiado cuando los dos conjuntos miden cantidades diferentes, como sucede en el gráfico de la página 59. Los gráficos de eje de *y* doble pueden confundir, ya que quien elabora el gráfico puede ajustar las escalas de los ejes para resaltar una idea particular. El ejemplo que utilizamos en el texto era un gráfico engañoso para representar las prácticas de Planned Parenthood.

Eje truncado. Comenzar un eje *x* o un eje *y* con un valor diferente del menor posible. A veces esto permite que el observador contemple más claramente la zona del gráfico

en la que están las observaciones. Pero cuando se usa de modo manipulativo puede distorsionar la realidad. El gráfico que mostramos en este glosario para ilustrar el gráfico de dispersión muestra dos ejes truncados eficazmente y no da lugar a una falsa impresión sobre los datos. El gráfico de las Noticias de la Fox que aparece en la página 48 proporciona una impresión falsa de los datos, tal como se muestra en la versión rediseñada que aparece en la página 49.

Exactitud. Lo próxima que se encuentra una cifra respecto a la cantidad verdadera de lo que se está midiendo. No debe confundirse con precisión.

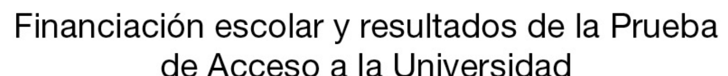
Extrapolación. Proceso mediante el que se realiza una suposición o inferencia sobre un valor (o valores) situados más allá del conjunto de valores observados.

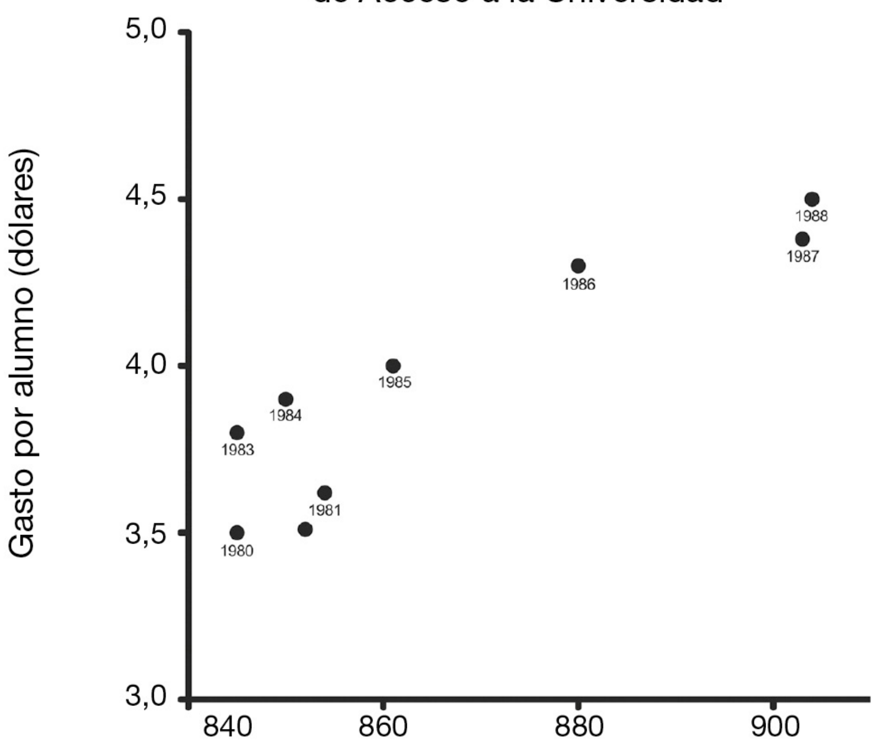
Falacia de afirmación de la consecuencia. Véase *conversión, error de*.

Falacia de la excepción. Error de razonamiento que tiene lugar cuando se realizan inferencias sobre un grupo basándose en el conocimiento de unos pocos individuos excepcionales.

Falacia ecológica. Error de razonamiento que tiene lugar cuando se realizan inferencias sobre un individuo basándose en datos agregados (como la media de un grupo).

GIGO[228]**.** *Garbage in, garbage out.*





Puntuación de la Prueba de Acceso a la Universidad

Gráfico de dispersión. Un tipo de gráfico que representa individualmente todos los puntos. Por ejemplo, en la página anterior se muestra un gráfico de dispersión de los datos presentados en la página 57.

Incidencia. La cantidad de casos nuevos (por ej., de una enfermedad) que se observan en un cierto periodo de tiempo.

Inductivo. Forma de razonamiento inferencial mediante el que un conjunto de observaciones conduce a un enunciado general.

Interpolación. Proceso para estimar cuál es el valor intermedio situado entre dos valores observados.

Inversión, error de. Tipo no válido de razonamiento deductivo de forma:

Si A, entonces B
No A
Por tanto, no B

Media. Una de las tres medidas de promedio (la tendencia central de un conjunto de observaciones). Se calcula sumando el total de las observaciones y dividiendo por la cantidad de observaciones. Es a lo que la gente se refiere cuando dicen «promedio». Los otros dos tipos de promedio son la mediana y la moda. Por ejemplo, para {1, 1, 2, 4, 5, 5} la media es (1 + 1 + 2 + 4 + 5 + 5) ÷ 6 = 3. Obsérvese que, al contrario que sucede con la moda, la media no es necesariamente un valor de la distribución original.

Mediana. Una de las tres medidas de promedio (la tendencia central de un conjunto de observaciones). Es el valor respecto al cual la mitad de las observaciones se sitúan por encima y la mitad por debajo. Cuando hay un número par de observaciones los estadísticos pueden emplear la media de las dos observaciones intermedias. Por ejemplo, para {10, 12, 16, 17, 20, 28, 32} la mediana es 17. Para {10, 12, 16, 20, 28, 32}, la mediana sería 18 (la *media* de los dos valores intermedios, 16 y 20).

Moda. Una de las tres medidas de promedio (la tendencia central de un conjunto de observaciones). Es el valor que sucede con más frecuencia en una distribución. Por ejemplo, para {100, 112, 112, 112, 119, 131, 142, 156, 199} la moda es 112. Algunas distribuciones son bimodales o multimodales, lo que significa que dos o más valores suceden un mismo número de veces.

Modus ponens. Un tipo válido de argumento deductivo de forma:

Si A, entonces B
A
Por tanto, B

Post hoc, ergo propter hoc (después de esto, por lo tanto debido a esto). Falacia lógica que surge de creer que porque algo (Y) sucede después de otra cosa (X), X *ha causado* Y. X e Y podrían estar correlacionados, pero eso no significa que exista una relación causal.

Precisión. Una medida del nivel de resolución de un número. El número 99 es preciso hasta 0 decimales y solo tiene la resolución del número entero más cercano. El número 909, 35 es preciso hasta el segundo decimal y su resolución es de una centésima de la unidad. Precisión no es lo mismo que exactitud: el segundo número es más preciso, pero si el valor verdadero es 909,00, el primer número es más exacto.

Prevalencia. La cantidad de casos existentes (por ej. de una enfermedad).

Promedio. Es un estadístico de resumen que se emplea para caracterizar cierta cantidad de observaciones. «Promedio» no es un término técnico y habitualmente se refiere a la *media,* pero también puede referirse a la *mediana* o a la *moda*.

Silogismo. Tipo de enunciado lógico en el que una conclusión se sigue necesariamente de las premisas.

Subdivisión. Descomponer un conjunto de observaciones en grupos menores. Resulta aceptable cuando hay heterogeneidad en los datos y el grupo mayor se compone de entidades que varían en alguna dimensión importante. Pero la subdivisión puede emplearse de modo engañoso para dar lugar a un gran número de grupos pequeños que no difieren sensiblemente en una variable de interés.

---

228 *N. del T.: Garbage in, garbage out.* Basura entra, basura sale; término utilizado en informática para indicar que el resultado de los cálculos de un ordenador es tan bueno como lo sean los datos que ingresan en él para realizar la computación.

AGRADECIMIENTOS

La inspiración para escribir este libro tiene su origen en *How to lie with Statistics* (Cómo mentir con la estadística), de Darrell Huff, un libro que he leído varias veces y que aprecio más con cada lectura. También he sido un fervoroso seguidor de *Damned Lies and Statistics* (Mentiras podridas y estadística), de Joel Best, y de *Naked Statistics* (La estadística al desnudo), de Charles Wheelan. Tengo una gran deuda con esos tres autores por su humor, su sabiduría y sus intuiciones y me gustaría que este libro ocupara un lugar junto a los suyos para cualquier persona que desee mejorar su comprensión del pensamiento crítico.

Mi agente, Sarah Chalfant de Wylie Agency es un sueño: cálida, atenta, alentadora e infatigable. Me considero un privilegiado por trabajar con ella y sus colegas de la TWA, Rebecca Nagel, Stephanie Derbyshire, Alba Ziegler-Bailey y Celia Kokoris.

También quiero mostrar mi agradecimiento a todos los que trabajan en Dutton/Penguin Random House. Stephen Morrow ha sido mi editor en cuatro libros y ha mejorado mucho cada uno de ellos (P<0,01). Sus orientaciones y apoyo han sido valiosos. Gracias a Adam O'Brien, LeeAnn Pemberton y Susan Schwartz. Me quito el sombrero ante Ben Sevier, Amanda Walker y Christine Ball por la cantidad de cosas que hacen para que los libros lleguen a los lectores que desean leerlos. Becky Maines ha sido una magnífica editora-de-más-de-una-revisión y he disfrutado de la amplitud y profundidad de su conocimiento y de sus esclarecedoras aportaciones.

Estoy agradecido a las siguientes personas por sus comentarios y contribuciones a los borradores del manuscrito: Joe Austerweil, Heather Bortfeld, Lew Goldberg y Jeffrey Mogil. Estoy en deuda respecto a párrafos específicos con David Eidelman, Charles Fuller, Charles Gale, Scott Grafton, Prabhat

Jha, Jeffrey Kimball, Howie Klein, Joseph Lawrence, Gretchen Lieb, Mike McGuire, Regina Nuzzo, Jim O'Donnell, James Randi, Jasper Rine, John Tierney y los muchos colegas de la American Statistical Associaton que han contribuido a revisar el libro y los ejemplos, en especial con Timothy Armistead, Edward K. Cheng, Gregg Gascon, Edward Gracely, Crystal S. Langlais, Stan Lazic, Dominic Lusinchi, Wendy Miervaldis, David P. Nichols, Morris Olitsky y Carla Zijlemaker. Mis estudiantes del Seminario de Menciones e investigación independiente de McGill me han ayudado con algunos ejemplos y me han permitido aclarar mis ideas. Karle-Philip Zamor, al igual que hizo con cuatro libros previos, me ha ayudado enormemente a preparar las ilustraciones y a resolver todo tipo de problemas técnicos, siempre con buen humor y gran competencia. Lindsay Fleming, mi ayudante de oficina, me ayudó a estructurar mi tiempo y a mantenerme centrado y me ha ayudado con las notas, el índice, la corrección de pruebas, la verificación de datos y muchos otros detalles del libro (y gracias a Eliot, Grace, Lua y Kennis Fleming por compartir su tiempo conmigo).

Agradezco a los siguientes lectores atentos que localizasen errores de la primera edición que en esta han sido corregidos: David L. Bitters, Andrew Brooks, Bruce Brown y Hap Miller.

Título original: *Weaponized Lies*

Edición en formato digital: 2019

Calle Juan Ignacio Luca de Tena, 15
28027 Madrid
www.alianzaeditorial.es

ISBN ebook: 978-84-9181-476-4

Conversión a formato digital: REGA

www.alianzaeditorial.es